CU6-H02963-10-P 001849

41 No. 1 (New Series No. 189) January 1981

ANALYSIS

Edited by
CHRISTOPHER KIRWAN

105
An 13

10

CONTENTS

note on Thomson's lamp 'paradox' — GEOFFREY C. BERRESFORD
low to travel faster than light? — ROBERT ELLIOT
he tree theory and isomorphism — CHARLES SAYWARD
asic and non-basic actions: 'same' or 'different'? — VIVIAN M. WEIL *and* IRVING THALBERG
eply to Weil and Thalberg — JENNIFER HORNSBY
n an objection to the synonymy principle of property identity — MICHAEL TYE 22
note on the substitutivity of notes — G. BAR-ELLI
An anti-essentialist note on substances — GRAEME FORBES
ewis and the problem of causal sufficiency — DAVID-HILLEL RUBEN
ocke on causation, compatibilism and Newcomb's problem — ANDRÉ GALLOIS
e inevitability of the state — DIANA T. MEYERS
ozick on Hart on the right to enforce — HILLEL STEINER
olating intrinsic value — EVA BODANSZKY *and* EARL CONEE
nperative inference—an addendum — VERA PEETZ
nagination, desire and prescription — H. M. ROBINSON
ill, paternalism, and slavery — JOHN D. HODSON
n a claim of Hare's — P. T. GEACH
a misunderstanding of Geach's — R. M. HARE

ASIL BLACKWELL · ALFRED STREET · OXFORD

Philosophy Documentation Center

The Philosopher's Index: An International Index to Philosophical Periodicals. An up-to-date quarterly index with abstracts from more than 300 international philosophy and interdisciplinary journals. An annual cumulative edition is also published. In addition, two retrospective indexes have been published, i.e., the *Retrospective Index to U.S. Publications From 1940* and the *Retrospective Index to non-U.S. English Language Publications From 1940.*

Philosophy Research Archives: A Bilingual Microfiche Journal of Philosophy, sponsored by the APA, the CPA, and the PDC, is the microfiche journal of philosophy which provides a means of publication that is free of many of the disadvantages and constraints of ordinary journal publication. In addition, the *Archives* accepts articles, translations, monographs, and bibliographies.

Teaching Philosophy. Explores ideas about teaching and learning philosophy by publishing articles, discussions, and reports on new courses, methods, materials and their impact on philosophy. Textbook and audio-visual reviews are regular features.

Directory of American Philosophers. A handbook of philosophy covering the United States and Canada with information on faculties of philosophy, names, addresses, and specialties of philosophers, data on colleges and universities, philosophical societies, centers, institutes, journals, publishers, and data on fellowships and assistantships (types and dollar amounts) available in philosophy. Statistics on the profession.

International Directory of Philosophy and Philosophers. A handbook of philosophy covering Europe, Central and South America, Asia, Africa, and Australia. This directory provides names, addresses, and specialties of philosophers as well as data on colleges, universities, philosophical societies, institutes, journals, and publishers of philosophical works.

Directory of Women in Philosophy. A handbook of women in philosophy, with information on fields of specialization, degree (date and institution), dissertation or publications, position, institution, dossier source, address, and telephone number.

Bibliographies of Famous Philosophers. Hardbound bibliographies have been published on Edmund Husserl, Jean-Paul Sartre, Alfred North-Whitehead, and Henri Bergson. Forthcoming bibliographies include works on Martin Heidegger, George Santayana, and a revised Sartre bibliography.

Dialog. A computerized search service which provides custom bibliographies with abstracts on any logical combination of philosophical topics.

Mailing Labels of approximately 10,000 philosophers in the United States and Canada are available as well as addresses by specialty, i.e., Ethics, Aesthetics, etc.

Journal Typesetting. Journals are economically typeset using computer driven photo-composition equipment. Books, ads, and other documents can also be done.

Bibliography of Philosophy. The PDC, the U.S. national center, selects books for review, and forwards the reviews to editors in Paris.

105 An 13

PHILOSOPHY DOCUMENTATION CENTER
Bowling Green State University
Bowling Green, Ohio 43403 USA

P1849

Vol
105
An 13

A NOTE ON THOMSON'S LAMP 'PARADOX'

By GEOFFREY C. BERRESFORD

IN 1954, J. F. Thomson proposed the following problem:

> There are certain reading lamps that have a button in the base. If the lamp is off and you press the button the lamp goes on, and if the lamp is on and you press the button the lamp goes off. So if the lamp was originally off, and you pressed the button an odd number of times, the lamp is on, and if you pressed the button an even number of times the lamp is off. Suppose now that the lamp is off, and I succeed in pressing the button an infinite number of times, perhaps making one jab in one minute, another jab in the next half-minute, and so on, ... After I have completed the whole infinite sequence of jabs, i.e. at the end of the two minutes, is the lamp on or off? It seems impossible to answer this question. It cannot be on, because I did not ever turn it on without at once turning it off. It cannot be off, because I did in the first place turn it on, and thereafter I never turned it off without at once turning it on. But the lamp must be either on or off. This is a contradiction.

Thomson was considering a lamp whose switch has exactly two states, on and off, and which can change instantly from one state to the other. We shall refer to this as an *abstract* switch. No mention is made of the state of the lamp at the transition times $t=1, 3/2, 7/4, 15/8 \ldots$, and presumably the lamp may be either on or off at these moments. For such an abstract switch, the state of the lamp at time $t=2$ minutes is indeed impossible to determine. The situation is analogous to a function f(t) on [0, 2], as in Figure 1, which takes on only two values, 0 (for on) and 1 (for off), and jumps discontinuously from one value to the other at the transition times $t=1, 3/2, 7/4, \ldots$.

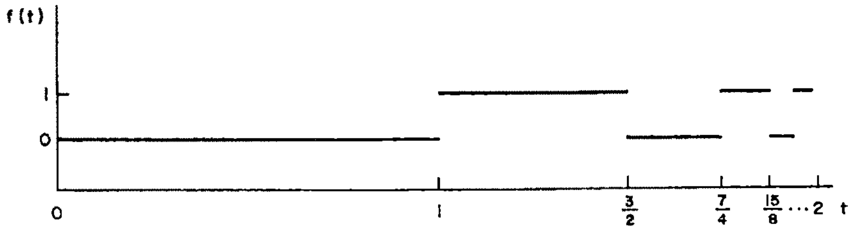


Figure 1

At each of the transition times it may be defined as zero or one. Since such a function may be chosen to be zero or one at time $t=2$, the final state of the lamp is indeed undecidable.

The situation, however, is entirely different if we consider any sort of *physical* switch. By a *physical* switch we shall mean one with two contacts, which are either apart, breaking the circuit, so that the lamp is off, or together, completing the circuit, so that the lamp is on. Such a switch actually has many possible states, corresponding to the distance between the contacts. Let us call $f(t)$ the distance between the contacts at time t minutes. We say the lamp is on or off at time t accordingly as $f(t)=0$ or $f(t)>0$. The lamp is then on when f takes values in a closed set, $\{0\}$, and off when f takes values in the complementary set, which is therefore open (as a subset of the range). Note the contrast with the symmetry of the states for the abstract switch. Since the contacts of a physical switch can move only continuously, $f(t)$ must be a continuous function on $[0, 2]$. Since the inverse of a continuous function takes open sets into open sets and closed sets into closed sets, the inverse image of the closed set $\{0\}$, $\{t|f(t)=0\}$, must be closed (in the domain $[0, 2]$). Note that this is the set of times at which the lamp is on. As this closed set contains values of t arbitrarily close to $t=2$, it must contain $t=2$, hence $f(2)=0$. We are thus led to the surprising conclusion that the lamp must be on at the conclusion of the switchings.

There was but one assumption made to obtain this result: the function $f(t)$ was assumed continuous on $[0, 2]$. The continuity was physically necessary, and the fact that the domain includes $t=2$ means the switch must be open or closed (not in a third state, 'broken') at time $t=2$. It is unnecessary to invoke any physical considerations such as velocities not exceeding the speed of light. It is enough to assume merely the continuity of motion.

There are, of course, continuous functions on $[0, 2]$ taking on zero and positive values in the required manner. Figure 2 is a graph of one.

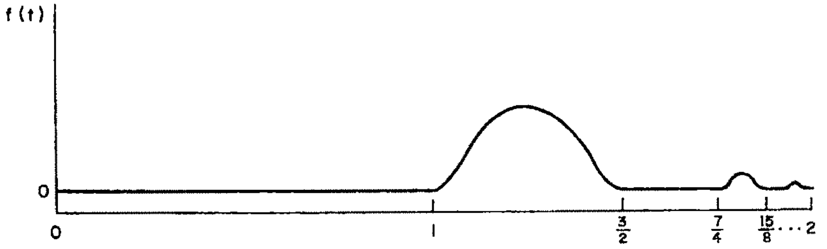


Figure 2

The continuity at $t=2$ requires, of course, that the successive maxima decrease to zero, showing again how the simple assumption of continuity leads to the conclusion that the lamp finishes on.

It is interesting to consider why, even for a physical switch, the question of the final state of the lamp is often considered unanswerable.

It would seem that, for any lamp which finishes on, one could construct a second lamp, possibly with the aid of a photoelectric cell, which would be off whenever the first is on and vice-versa. The second lamp would then finish off, contradicting the result. A moment's thought, however, will convince one that this is not the case. The current in the second lamp must ultimately be controlled by a switch able to open and close as rapidly as that of the first lamp. However, we have seen that any switch able to perform in the necessary manner will be closed at the conclusion of the switchings, leaving the lamp on, as before.

Long Island University

© Geoffrey C. Berresford 1981

REFERENCES

J. F. Thomson, Analysis 15, 1, 1954, pp. 1–10.
A. Grünbaum, *Modern Science and Zeno's Paradoxes* (Wesleyan University Press, Middletown, Conn., 1967).
A. Grünbaum, Science 159, 26 (Jan. 26, 1968), pp. 396–406.
W. C. Salmon, ed., *Zeno's Paradoxes* (Bobbs-Merrill, Indianapolis and New York, 1970). This contains a revised version of Thomson's article cited above.

105 An 13

P 1849

HOW TO TRAVEL FASTER THAN LIGHT?

By ROBERT ELLIOT

THE inter-stellar beam-transport system has been working smoothly for years now. All one has to do is attend the departure centre at the appropriate time. The rest is simple. One's biochemical details are recorded and one's body is then disintegrated. The recorded biochemical blueprint is electro-magnetically transmitted to the arrival centre at the desired destination. The traveller is then re-embodied in accordance with the blueprint out of the appropriate elements from the element bank. However some people were dubious about beam-transport in the early days. What worried them was the fact that beam-transport involved disintegration of the body. They found it hard to accept that a person could survive bodily discontinuity. However beam-transport does nothing to interfere with psychological continuity and cohesiveness. Travellers leave and arrive with the same psychological characteristics, and there is, after all, a direct causal chain which links a traveller's final departure stage with his initial arrival stage. As the philosophers showed long ago these conditions are sufficient to ensure personal survival.[1]

Beam transport does have its limitations. Despite the fact that it enables travel at the speed of light it still involves enormously lengthy transit times. The problems this causes for social and personal relationships have been well documented. The solution though, is quite simple. Let me explain.

Some time ago I heard a story about Argon, who wanted to go to Alpha Centauri. He attended the departure centre on Earth and his biochemical details were recorded and his body disintegrated. However, something quite unprecedented happened. A machine malfunction resulted in the destruction of Argon's blueprint. Miraculously Argon survived, since something equally unprecedented had happened later on Alpha Centauri. A machine malfunction resulted in the embodiment of an individual conforming to Argon's blueprint. The two events were not causally connected, and Argon found himself on Alpha Centauri four years too early.

In fact I heard this story at the Alpha Centauri departure-centre, and was thus able to put it to good use. I knew that I would have to return to Alpha Centauri in ten years' time. This was not a pleasant prospect. In fact I would spend the greater part of the next ten years in transit. Taking inspiration from Argon's "mishap", I spoke to the Controller at the Alpha Centauri departure-centre and persuaded him to keep a copy of my blueprint. The idea was that in ten years' time the Controller would see to it that I was embodied on Alpha Centauri, just prior to which time I was to attend the departure centre on Earth where my body would be disintegrated. No biochemical blueprint would need to be transmitted. Of course the trick can be worked in both directions and I therefore made the appropriate arrangements with the Controller at Earth-centre to ensure my earliest possible return.

---

[1] See, for example, Derek Parfit, 'Personal Identity', *Philosophical Review*, 1971, and John Perry, 'The Importance of Being Identical', in A. O. Rorty (ed.) *The Identities of Persons*, Berkeley, 1976. The arguments contained in these papers are decisive in establishing a theory of personal survival which should allay any fears one might have concerning beam-transport.

As you by now realise I have made my return visit to Alpha Centauri and I am back on Earth once more. It is evident to me that one can travel considerably faster than the speed of light if one makes judicious arrangements concerning embodiments of one's blueprint. I must confess that I was a little worried at first about surviving, but on reflection I dismissed my doubts. After all the situation in faster-than-light travel is much the same as it is in the case of conventional beam-transport. The person arriving on Alpha Centauri is psychologically continuous and cohesive with the person who departs from Earth. There is also a causal explanation of this, though it is not in terms of any direct causal link between the initial arrival-stages of the traveller and his final departure-stages. The relevant cause of my arrival considerably pre-dates the events which take place at departure.

This suggests something interesting, namely that those theories of personal identity which allayed our misgivings about beam-transport are stronger than need be. The causal continuity requirement built into them (Parfit p. 11 n.) is otiose. Strictly speaking there is no causal continuity in faster-than-light travel since not every person-stage is causally connected to its adjacent person-stages. That, however, is not cause for alarm since causal continuity offers no logical guarantees not equally available to those who wish to use the faster means of travel. Certainly it no more rules out the possibility of traveller duplication;[2] that is, the possibility that the traveller is embodied twice over. Whether or not someone arrives, irrespective of which method of travel they use, will depend on the adequacy of such things as timetabling arrangements, transmission apparatus, technician competence and the like. Empirically speaking there may well be less risks with the faster method. Furthermore when we reconsider Argon's case it emerges that even the more roundabout causal connection plays no logically significant role in ensuring survival. The very same point can be made about traveller duplication. The only reason for not relying on Argon's method of travel is the extreme improbability of it ever succeeding. His survival was, quite literally, an accident.

---

[2] Critics of beam-transport, inspired by Bernard Williams' *Problems of the Self*, Cambridge, 1972 (see page 24), have appealed to this possibility to support their belief that beam-transporting effectively kills the would-be traveller.

Given the great advantages of faster-than-light travel it seems rational to use it. Any doubts we have about whether we would survive, assuming there are no technical mishaps or human errors, are groundless. Since the causal continuity requirement offers no logical guarantees any doubts we have about faster-than-light travel are equally doubts about beam-transport. Both methods stand or fall together. Moreover we do know that travellers survive beam-transporting. Don't we?

University of Queensland © ROBERT ELLIOT 1981

THE TREE THEORY AND ISOMORPHISM

By CHARLES SAYWARD

A MAIN thesis of Fred Sommers' type theory, as set forth in [4], is that an isomorphism exists between any natural language and the categories discriminated by that language. Here I give an explanation of what this claim comes to. And then I argue that, so understood, the claim is incompatible with Zermelo-Fraenkel set theory. Finally, I argue against trying to salvage the isomorphism thesis by appealing to some other set theory.

I

Following Suppes ([5], p. 262) I shall say, for any sets A, B and relations R, S: $\langle A, \mathrm{R}\rangle$ is isomorphic to $\langle B, S\rangle$ iff there exists a function f such that

(1) f is one-one;
(2) the domain of $f = A$ and the range of $f = B$;
(3) if $x, y \in A$, then $\langle x, y\rangle \in \mathrm{R}$ iff $\langle f(x), f(y)\rangle \in S$.

To make Sommers' thesis intelligible we must determine what he would take the requisite A, B, R, S and f to be.

There are different sets of assignments that are attributable to Sommers. These turn out to be equivalent. The one I choose seems to me to be the most elegant and to require the least elaboration.

To see what is to be assigned to $\langle A, \mathrm{R}\rangle$ consider the following equivalence relation:

P and Q are significantly predicable in L of the same individuals.

This relation partitions the monadic predicates of a language into equivalence classes. These classes are what Sommers calls 'A-types':

> A set X of predicates of a language L is an A-type for L iff, for some predicate P and any predicate Q of L, $Q \varepsilon X$ iff P and Q are significantly predicable in L of exactly the same things.

The relation R is defined on the set of A-types:

> If X and Y are A-types, then X bears R to Y iff, for all $P \varepsilon X$ and $Q \varepsilon Y$, P significantly applies to whatever Q significantly applies to.

For the $\langle A, \mathrm{R}\rangle$-side of the isomorphism, let A be the set of all A-types for L, and R be the relation just spelled out.

The concept of a category is crucial to understanding how to take the $\langle B, \mathrm{S}\rangle$-side. A category is a set of individuals of which a particular predicate is significantly predicable.

> X is a category for L iff, for some predicate P of L and any individual x, $x \varepsilon X$ iff P is significantly predicable of x in L.

$\langle B, S\rangle$, then, is to be interpreted in this way:

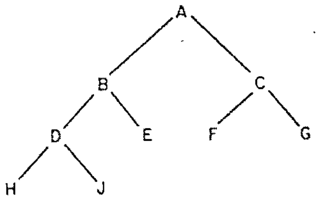
> B: the set of categories for L;
> S: proper inclusion.

The function f, which Sommers has in mind, can be set out as follows:

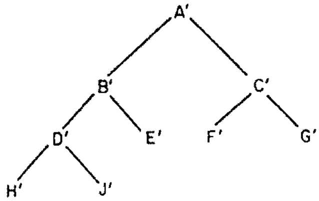
> $f(X) = Y$ iff X is an A-type for L *and* Y is a category for L *and*, for some predicate P of L in X and for any individual x, $x \varepsilon Y$ iff P significantly applies to x in L.

Since any two predicates in a given A-type significantly apply to the same individuals, we could have put 'for each predicate P' for 'for some predicate P' in the above characterization. The substantive issue before us is whether the function f satisfies (1)—(3).

For Sommers the pair $\langle A, \mathrm{R}\rangle$ comprise a structure, which may be graphically represented by a tree. Suppose the monadic predicates of L fall into A-types A, B, . . . , J. A possible representation of these A-types is this:



According to the isomorphism thesis, if the tree just given represents the $\mathcal{A}$-types for L, the following tree represents the categories for L:



In the case of the language tree the nodes are $\mathcal{A}$-types; if one $\mathcal{A}$-type is connected to a lower one by a downward path, the higher one bears R to the lower one; any two nodes on the tree are disjoint.

In the case of the ontological tree the nodes are categories; downward paths mean proper inclusion; two sets not connected by a downward path are disjoint.

II

If the axioms of Zermelo-Fraenkel set theory, ZF, are true, the set of $\mathcal{A}$-types for English is not isomorphic to the set of categories for English in the way Sommers claims the two are isomorphic.

Consider the predicate '$x=x$'. This is true of everything and so it is significantly applicable to everything. It is, accordingly, an example of what Sommers calls an 'apex term'. So the $\mathcal{A}$-type it belongs to is at the top node of the language tree.

There is no corresponding category—at least, not if one accepts the axioms of ZF. According to that theory, there is no universal set. And for there to be a category corresponding to the top $\mathcal{A}$-type there would have to be a universal set. So we have an $\mathcal{A}$-type with no corresponding category, in which case Sommers' isomorphism thesis is false.

III

A *set* and a *proper class* are, respectively, a class which is a member of some class and a class which is a member of no class. Call a theory about classes a set theory if it quantifies over sets only; call it a proper class theory if it quantifies over both sets and proper classes. Consider the proper class theory, VN, expounded in [2], utilizing ideas of von Neumann and others. According to VN '$x=x$' has an extension. This extension, of course, is a proper class. This points to a suggestion for salvaging Sommers' isomorphism thesis: understand *A*-types and categories to be classes. The top node of the ontological tree is a proper class.

I object to this alternative on the grounds that there are good reasons for doubting whether VN is a coherent theory.

The standard characterization of an interpretation of a first order language *L* is familiar. An interpretation consists of a non-empty class, called the domain of discourse, plus a function that assigns the usual things to the predicates and individual constants of *L*. When one goes to the definition of truth under an interpretation, it is plain that the domain of discourse is the class of those entities which are values of the variables of *L* under that interpretation.

Consider now a deductive theory *T* formalized in *L* whose sentences are true or false antecedent to any semantic account of *T*. Such an account, if adequate, defines a predicate whose extension comprises the sentences true in *T*. Looked at as an uninterpreted formal language *T* has all kinds of interpretations. A correct semantic account will describe that interpretation which matches *T*'s antecedent meaning as nearly as possible. Once that interpretation is specified the truth predicate for *T* is defined by the clauses that define truth under that interpretation.

This clearly will not do where *T* is VN. For the method requires describing the theory's domain of discourse, which is the class of those entities which are values of the theory's variables. But among the values of variables of VN are proper classes. Hence VN has no domain of discourse.

An alternative for defining truth for first order theories is a Tarski-style definition. This does not require describing the theory's domain of discourse.

But crucial to a Tarski-style definition of truth for a first order theory is that the metalanguage in which the definition is formulated quantify over all infinite sequences of objects over which the object language quantifies. In the case of VN these objects include proper classes. And proper classes cannot be elements of sequences, which are classes of a certain sort. It is owing to this fact that a Tarski-style definition of truth for VN cannot be adequate.

Neither the domain of discourse approach nor the Tarski-style approach provide a semantics for VN. Is there a third alternative? I know of none. Every variant of these approaches I can think of in the end requires that all the objects VN quantifies over be elements of a set theoretic object—either a function, a sequence or some other type of class—which the metalanguage quantifies over. Until this problem is ironed out VN has to be regarded as a dubious account of classes. Right now the reasonable belief is that all classes are sets.

IV

Other alternatives to ZF are set theories which assert the existence of a universal *set*. An example is the theory NF articulated by Quine in [3]. Could not Sommers simply opt for such a set theory?

A set theory that asserts the existence of a universal set must deny ZF's axiom of foundation. Fraenkel has the following to say about this axiom:

> Examples [of extraordinary sets] are a set that contains itself as a member and a set s, which together with a set t, satisfies $s \varepsilon t$ and $t \varepsilon s$. In order to exclude such (certainly unnecessary and undesirable) sets, von Neumann (1925) and, later, others proposed the addition of an axiom (of "foundation" or "restriction") of the following type:
>
> . . . Every nonempty set s contains a member t such that s and t have no common member ([1], p. 424).

The universal set is an extraordinary set. So if Sommers (or a defender of the isomorphism thesis) defends that thesis by adopting a set theory that countenances the universal set, he must take into account the motivation Fraenkel gives for the axiom that denies the existence of extraordinary sets. This is that such sets are 'unnecessary and undesirable'. Why does Fraenkel say this?

As far as I know, there is no commonly accepted body of knowledge which depends upon there being extraordinary sets. For example, there exists no commonly accepted result in mathematics which requires extraordinary sets. It is such considerations, I suppose, which Fraenkel has in mind when he says extraordinary sets are unnecessary.

But now Occam's maxim ('Don't multiply entities beyond necessity') is relevant. If extraordinary sets are not necessary for deriving wanted results, then that is an excellent reason for regarding them as undesirable. This is reinforced by the fact that the existence of a set K={. . . K . . .} runs counter to many people's intuitions about sets.

A further fact which makes the universal set undesirable is this. Consider this principle:

(4) If S is a set and P is a meaningful predicate there is a set S_p consisting of just those elements x of S which satisfy P.

If P is true of all elements of S then $S_p = S$; if P is true of no elements of S, S_p is the null set. In any other cases S_p is a proper subset of S. It is hard to see what fault could be found with (4). But if S is the universal set and P is 'is not a member of itself', then a contradiction results. So if there is a universal set, either (4) is false or 'is not a member of itself' is meaningless. But the latter course cannot plausibly be taken since

(5) The universal set ε the universal set

is true if there is a universal set. This means 'is a member of itself' must be a meaningful predicate. How could this be so if 'is not a member of itself' were meaningless? It looks as if a defender of the universal set has to reject (4). It would be interesting to see what argument could be given in defence of such a move.

V

I have seen telling refutations of Sommers' own arguments for his tree theory. But I have not seen a refutation of the theory itself. It remains an ingenious and plausible speculation about natural languages. But the isomorphism thesis is another matter. It requires positing proper classes or extraordinary sets. Neither course is plausible.

The University of Nebraska—Lincoln © CHARLES SAYWARD 1981

REFERENCES

[1] Abraham Fraenkel, 'Set Theory', *The Encyclopedia of Philosophy*, Vol. 7, edited by Pau Edwards, Collier-McMillan (1967), pp. 420–7.
[2] J. Donald Monk, *Introduction to Set Theory*, McGraw-Hill, 1969.
[3] W. V. O. Quine, 'New Foundations for Mathematical Logic', *From a Logical Point of View*, Harvard University Press, 1953.
[4] Fred Sommers, 'Types and Ontology', *Philosophical Review*, Vol. 72 (1963), pp. 327–63.
[5] Patrick Suppes, *Introduction to Logic*, Van Nostrand, 1957.
[6] Alfred Tarski, 'The Concept of Truth in Formalized Languages', *Logic, Semantics and Metamathematics*, Clarendon Press, 1956.

BASIC AND NON-BASIC ACTIONS: 'SAME' OR 'DIFFERENT'?

By VIVIAN M. WEIL *and* IRVING THALBERG

JENNIFER HORNSBY (ANALYSIS 39.4, October 1979) criticizes a method of act-individuation originally developed in our joint essay, 'Elements of Basic Action' (*Philosophia* 4.1, 1974), and refined in Thalberg's *Perception, Emotion and Action* (Blackwell, 1977). We believe the issues and Hornsby's reasoning merit further ventilation.

Our example should evoke the issues. In faraway Acirema, around time t_1, the Prime Minister (henceforth abbreviated 'PM') becomes convinced that neighbouring Dauphinia (henceforth 'D') is all set to unleash its nuclear missiles. Acirema's own atomic weapons are ready to careen pre-emptively towards D. All it takes is a signal from the flashing red buzzer on PM's desk. So he swings into action: he crooks a finger, thereby pushing the buzzer, dispatching his missiles, and erasing D.

What would rival act-individuationists report? Multipliers or pluralists like Goldman would catalogue an array of deeds—each presumably as distinct from the other as PM's finger-bending at t_1 is from, say, his grimacing at t_1. Reductive unifiers, led by Davidson, would record but one action—PM's 'primitive' or basic action of finger-bending—and several misleading specifications of that action in terms of its later effects. On this view, PM's digital movement and his flattening of D share all their properties, including spatio-temporal dimensions. Thus Davidson would say it took no longer for PM to level D than to crook his finger. Our own non-reductive unifying proposal was to tally but a single action. However, you may delineate what is done in various non-coextensive terms: pressing a buzzer, releasing the missiles, pulverizing D—or simply moving a finger. We think the wider-ranging non-basic deeds cannot be said to have occurred until later than t_1, whenever the requisite additional event-components either join or succeed the muscular contractions, neural discharges and other sub-cutaneous happenings which make up PM's basic digital gesture. For us, those sub-occurrences which constitute his finger-flexing are also components of his non-basic deeds. Therefore his non-basic buzzer-pushing, missile-releasing and so on are not what might be called 'countably different' actions from his finger-bending. We will mark this familiar notion of difference with the subscript 'c'. Evidently PM's scowling and finger-bending are different$_c$ actions. If we are individuating objects, here is a parallel. Suppose you trade in your bicycle for a motorbike. Then you have acquired a different$_c$ means of transportation. We could put both vehicles side by side and count them. But if you merely attach a small engine to your trusty old bicycle, then what you ride now is not a different$_{c}$ two-wheeler.

Hornsby disdains this approach to individuation. She prefers to ask which deeds are either 'identical' ('the same') or else 'different' in the sense of being non-identical ('not the same'). The subscript 'ni' can serve to indicate this kind of difference. Broadly speaking, Hornsby thinks you can decide

> whether *a* is the same action as *b*, without . . . [asking] *How many actions?* . . . [I] dentity questions come first; . . . and . . . divide . . . Davidson from Goldman before any counting questions are allowed . . .
>
> . . . Take any pair of descriptions in the series 'his moving his finger' . . . 'his destroying [D]', construct an identity statement using [them] . . . and you will have a statement that . . . Davidson hold[s] true, Goldman false (p. 195).

We are unsure what sort of priority or pre-eminence Hornsby attributes to identity questions. But her fixation upon 'the same or different$_{ni}$?' may have prevented her from understanding our non-reductive unifying scheme. In the freshly coined terminology, our claim was that PM's basic and non-basic actions are not different$_{c}$. Hornsby appears to believe this must mean 'not different$_{ni}$' and says, 'it sound[s] as though . . . the actions [are taken] to be the same' (Hornsby, p. 196). However, '*a* does not differ$_{c}$ from *b*' hardly implies '*a* does not differ$_{ni}$ from *b*', *viz.*, '*a* is identical with *b*'. Your old two-wheeler, now equipped with an engine, is not a different$_{c}$ vehicle. We cannot put the contraption you used to have next to the one you presently ride. But a vehicle which is power-driven cannot be identical, in a straightforwardly Leibnizian sense, with a motorless one.

Hornsby seems to misunderstand this model for the relationship of non-basic to basic action. In her view, our contention is

> that the very event that is [PM's] moving his finger becomes his destroying [D] by addition to it of other events—just as a tadpole may grow legs and thereby become a thing with legs . . . [But] an event of missiles' blasting off is [never] part of a man's moving his finger—as the legs of a tadpole are at some time part of it (p. 196).

Hornsby's tadpole is a godsend. Our dictionary says that 'tadpole' probably once meant 'toad that is all head'. Nowadays tadpoles are defined as 'larvae of the order Salienta (the frogs and toads)'. However, you can use the term narrowly, for 'larvae which have lost . . . external gills but not yet acquired . . . forelimbs'. Strictly speaking, then, as soon as Hornsby's larva grows legs, it ceases to be a tadpole. Linguistic niceties aside, it is now 'a thing with legs', which it was not earlier. We seem to have different$_{ni}$—though not different$_{c}$—creatures on our hands. Of course the plain truth is that we have one evolving organism throughout. As for PM, it makes no sense to suppose that the 'missiles' blasting off' is made 'part' of his 'moving his finger', thereby transforming the finger movement into his 'destroying of [D]'. Only the sequence initiated by his digital movement goes on; it is one evolving sequence.

We urge Hornsby to distinguish between the accretion of new event-components to an expanding series, and somehow enlarging a component. She remarks:

> Continuants take up space . . . and . . . acquire spatial parts as they persist . . . Events take up time . . . and though their own parts may continue to happen as time progresses, we cannot make sense of events acquiring other events as parts of themselves . . . [W]e cannot say of an event *a* that it acquires *e* as a part, since the event $(a+e)$ that has *e* as a part must be different from *a* (p. 199).

We won't investigate what Hornsby means by an event's 'own parts'. This passage intrigues us because it elaborates her Pickwickean notion of 'adding on'. Obviously you cannot make 'other events . . . parts of' PM's basic finger-bending. What could it mean to say of this *a* 'that it acquires *e* [the ruin of D] as a part'? Yet it is meaningful and true to say that *e* is part of $a+e$. This latter sequence is indeed $\text{different}_{\text{ni}}$ from *a*, but not at all $\text{different}_{\text{c}}$ from *a*.

We hope this clears up Hornsby's interpretation of our non-reductive unifying method of individuating behaviour. Now we should confront her more or less direct arguments against our scheme. These are mainly attempts to defend the reductive claim that PM's non-basic buzzer-pressing and flattening of D are simply identical with his basic finger-flexing. Hornsby thinks no subsequent happening *e* must take place, and join up with *a*, for these non-basic deeds to occur. Therefore she must challenge a familiar assumption about how such non-basic goings-on are reported. The assumption is 'that the non-basic action cannot be said to have happened until later than the basic one' (Hornsby, p. 197). The obvious reason why both pluralists and non-reductive unifiers believe this 'cannot be said' is that the non-basic action itself seems not to be over until various effects of the basic deed have unfolded. For instance, PM would not seem to have blasted D off the map until nuclear explosions have occurred there at t_{1+n}, as a result of his earlier finger-bending in Acirema at t_1. Hornsby must explain why these correlative beliefs about what must take place and what can be said are mistaken.

Her attempted proof bewildered us somewhat, so we'll quote it lavishly. First she seems to hold that if you want to know when some non-basic feat has occurred, you date it by referring to the effect of a basic act. Thus:

> If somebody calls an action . . . a destroying of [D], then . . . he wants to know when . . . [D] was destroyed . . . And of course the various effects of a single [basic] action . . . occur at different times . . .

> . . . But even if one agrees that 'He has destroyed [D]' would be false if spoken before the . . . devastation, it is still possible to query whether that proves that his action of destroying [D] has not occurred then [*viz.*, before the devastation] . . . ' He has destroyed [D]' is true of him only if he has the property that anyone who has caused [D's] destruction has. But this is a property that [PM] does not come to have until something he does *has* caused the destruction (pp. 197f.).

Hornsby does not tell us how you can accept this analysis of the truth-conditions for 'He has destroyed D', and still claim that PM's non-basic action of destroying D has occurred prior to the fatal bomb blasts. Instead of dealing with that hitch, she seems to change her mind, and dismiss the foregoing doctrines about when non-basic actions are terminated. Apparently she believes such doctrines engender an absurdity, implying that PM

> comes to have destroyed [D], and his action comes to be a destroying of [D], at a time later than the occurrence of his action that is his destroying [D]. On this account . . . the tensed sentences are a distraction: they direct our attention away from the crucial question of when the action occurred, and towards the different question of when it can first be said to have certain properties (p. 198).

We suspect the absurdity results only from Hornsby's question-begging identification of 'his action that is his destroying [D]' with PM's finger-movement. She must have assumed an answer to 'the crucial question of when the action occurred': at the time PM finished moving his finger. Another snag in the above reasoning is that she seems to admit that PM's 'action' (finger-bending) does not yet possess the 'property' it needs to qualify as 'a destroying of [D]'. How then 'is' it already 'his destroying [D]'? She hasn't converted us to her reductive view.

But Hornsby's most appealing pro-reductive argument is still to come. More important, it does not seem to depend upon any of the contentions we have belaboured. To facilitate discussion, we have numbered each of her key statements. Hornsby declares it 'reasonable to think'

> (1) that an action that takes a stretch of time to occur is occurring at any moment in that stretch, and . . . (2) that if *a* is [an] agent, then any moment at which his action is occurring is a moment when *a* is doing something. (3) If . . . [PM's] destroying [D] goes on occurring at a later time than that at which he moves his finger, then . . . he must be doing something after he moves his finger. (4) But suppose . . . he dropped dead the moment his finger left the button. (5) [He still] did destroy [D]. (6) But . . . he did nothing after he pressed the button . . . (7) So his action of destroying [D] cannot have been occurring after . . . he moved his finger (p.198).

Finally, using component terminology, Hornsby proclaims (8) that no event which is 'someone's doing something' may contain 'a part that is not his doing anything' (p. 199). Throughout her reasoning, the notions of uninterrupted 'occurring' and 'doing' bear a heavy dialectical burden. For the sake of argument, however, we'll pretend to understand what such relentless goings-on are like. We want to focus on the statements in which these notions figure.

Her premise (1) strikes us as plainly false. Singing may go on all afternoon in the concert hall. But there will be pauses and intermissions during which vocalizing stops. Hornsby's (2) is equally dubious. While someone is doing her laundry, she may be merely standing by, listening to the operations of the machine as it comes to a finish. Even basic actions need not go through without gaps. The movements of Tai Chi exercises, for instance, are marked by periods of complete repose. If you generalize statement (3), you will find exceptions galore. Consider the operator of a knitting machine. She selects yarns and a pattern, then starts her machine. An alarm will ring if the machine needs tending. The woman might actively watch, and have to intervene occasionally; but she may doze throughout.

Now it is consistent with our objections against Hornsby's statements (1), (2), and their logical consequence (3), that we should accept (4)–(6). Our parallel to Hornsby's lugubrious case is that the woman's knitting machine is so well designed that it usually requires no supervision once she gets it going. If she then dies, she has nevertheless made the sweater that ultimately emerges—just as much as if she had lived to admire it.

This brings us to Hornsby's reductive conclusion (7). Shall we maintain that the actions of our late protagonists *have* been occurring after their death? We think the pressure to say that comes from Hornsby's discredited assumptions (1)–(3). Therefore we withhold any categorical verdict on (7), as to whether PM's action of destroying D kept on 'occurring' after his demise.

Instead we should remind ourselves why we consider this long series of incidents PM's action. The buzzer did not activate itself, nor was it set off by a short circuit. Acirema's missiles did not leave their launching pads spontaneously, or just happen to fall upon D. These goings-on were deliberately initiated by PM—as much as his finger-bending was.

Such reminders bear directly on Hornsby's conclusion (7). PM's goal in triggering the events we listed was the annihilation of D. Neither this culminating event nor the intermediate episodes took place until after he moved his finger. Since they were what he was trying to bring about, it seems arbitrary to deny that they were all part of the sequence forming his action. Judith Jarvis Thomson illustrates this point with her story of a woman cleaning the kitchen floor with Super-Stuff, which gradually disintegrates the dirt, leaving no residue to be swept up

(*Acts and Other Events*, Cornell, 1977, pp. 55ff.). Super-Stuff's eradicating the dirt is an essential part of her cleaning the kitchen floor. Even if she passes away after applying Super-Stuff, Thomson argues, 'her cleaning of her kitchen floor is not over until Super-Stuff has . . . disintegrated her dirt. For it is not over until the floor is *clean*' (p. 57). Reductive unifiers like Hornsby cannot hope to convert us without proving that this line of reasoning is unsound.

A final comment on Hornsby's statement (8) may diminish their zeal. Why must 'someone's doing something' contain no 'part that is not his doing something'? Flowers are made up of petals, stamens, pistils and stems: are these mini-flowers? Are the gears, springs and hands of a wristwatch themselves small-scale timepieces—and so on *ad infinitum*? Think about those very episodes which reductive unifiers call our only real actions—our basic or 'primitive' bodily movements: these seem to consist of muscular contractions, tendon stretchings, and signals from our local efferent nerve endings which control our muscles and tendons. In turn you can analyse these muscular and neural occurrences in electro-chemical terms. However you delineate these microscopic processes, they would seem to be the event-parts which make up any basic deed we perform. Yet they do not appear to be further actions of ours—instances of 'doing something'.

The upshot is simple: if reductive unifiers will concede that these *bona fide* primitive actions have event-parts which are not actions, then they should allow our long-range deeds to contain non-actions like the missiles' flight and the resulting havoc. You do not have to assert that PM's action—of bending his finger or destroying D—is still 'occurring', much less that he is still acting. All you have to say is that his action—the sequence of events he inaugurated by crooking his finger—is not completed until the missiles explode. This sounds more sensible than reporting that he has already flattened D the moment he has bent his finger, because these actions are 'the same'.[1]

© VIVIAN M. WEIL *and* IRVING THALBERG 1981

Center for the Study of Ethics in the Professions,
Illinois Institute of Technology, and
University of Illinois at Chicago Circle

---

[1] We thank the Editor for comments on an earlier draft.

REPLY TO WEIL AND THALBERG

By JENNIFER HORNSBY

'ACTIONS AND IDENTITIES' (ANALYSIS 39.4, October 1979) did not propose that we should report the Prime Minister as 'having already flattened Dauphinia the moment he has bent his finger'. It explained, in the light of general considerations about tense consistent with his bending of his finger's being his flattening of Dauphinia, why we should not make that report. This is one among a number of misrepresentations by Weil and Thalberg (ANALYSIS, this issue). Perhaps they misrepresent me because they think I misunderstood their non-reductive unifying scheme and disdained their approach to individuation. But I think that I did understand the scheme, and only maintained that the approach leads to falsehoods.

Even for someone fixated on *same or different*$_{ni}$*?*, it is easy to understand that if one uses 'different$_c$' in such a way that no pair of events falls inside its extension if they have components in common, then a, $a+e$, $a+e+f$ are not different$_c$. My objection was not to that stipulation, but to the claim that the action descriptions 'his moving his finger', 'his pressing the button', 'his destroying Dauphinia' can stand for things related as events in the series a, $a+e$, $a+e+f$. Thalberg had not only made that claim. He was also committed to the view that such descriptions stand for things that are not different$_{ni}$. Or so I thought. And if one puts together the claim and the commitment, one arrives at something that Weil and Thalberg say I said was Thalberg's contention and that they seem now to find as problematic as I did—that an event can acquire its *own* parts. For e surely is one of $a+e$'s parts; but then if a is the same as $a+e$, e is one of a's parts, and if a becomes $a+e$ by e's accretion (if 'it takes on broader dimensions with time'), a has acquired one of its own parts.

That was how I saw the state of the argument. And one extra piece of evidence can be given for the commitment I have just alleged. Thalberg apparently concedes in his book that an action of killing someone is an action of moving the body which causes his death.[1] So it seems that he cannot help but allow that, in the story he reminds us of, the action of destroying a city was an action of the Prime Minister's moving his finger which resulted in a city's destruction. But this is to say that there was an action which was his destroying of the city *and* which was (the same as, not different$_{ni}$ from) his moving his finger. Well, Thalberg would evidently wish to be free of this commitment, because he now asserts its contradictory quite explicitly. (Taking the destroying and the moving to be related as $a+e$ and a, he says at p. 14 that $a+e$ is different$_{ni}$ from a.) But then I would ask whether he thinks that he can after all withhold the very natural claims about 'kill' and other causative verbs, or whether he thinks that he can break the ties between identity and predication which enable us to pass from something's being ϕ and ψ to there being something which is ϕ and is the same as something which is ψ. The Weil-Thalberg scheme can only survive if it lets go of one or other of these things.

---

[1] This is actually a weakening of the claim conceded. *Perception, Emotion and Action* (Blackwell, Oxford, 1977) p. 102. Page references which follow in the text will be either to Weil and Thalberg's paper, above, or to my 'Actions and Identities'.

II

I was anxious to show that there is no difficulty in combining a view of action identities like Davidson's with the actual tense behaviour of English verbs. This objective has effectively been obscured by Weil and Thalberg's decision to conflate, by running together in an extended but much edited quotation, two claims that I carefully distinguished, and then to misreport both. As to the first, I did not say that you date *the occurrence* of a non-basic feat by reference to its effects: what I said (second complete paragraph on p. 197) was that you answer questions about when events *take effect* by reference to effects. As to the second, I did not challenge the familiar assumption that it cannot be said that the non-basic action has happened until a time later than that at which it can be said that the basic action has happened: what I challenged (as the third sentence they quote from my text of p. 197 makes plain) was the presumption that there is a valid passage from that familiar assumption to the claim that basic and non-basic actions occur at different times. Thus what Weil and Thalberg say that I find absurd—sc. that his action comes to be (gets the status of) a destroying of the city at a time later than its occurrence—I believe in fact to be true. (I cited Jonathan Bennett's 'Shooting, Killing and Dying' in support of that belief. *Canadian Journal of Philosophy* 1972.) And the premiss of Judith Jarvis Thomson which they spring on me is just an application of the step that I hoped to have cast in doubt. (If we cannot assume that, from the fact that 'x has ϕ-d' was first true at t, it follows that t was the last moment of x's ϕ-ing, then we cannot assume that 'her cleaning of the kitchen floor is not over until Super Stuff has disintegrated the dirt'. That assumption is no more secure than the assumption that the Prime Minister has flattened Dauphinia as soon as his action is finished.)

III

Since Weil and Thalberg think that all of the pressure to resist the conclusion of my positive argument for action identities comes from the premisses that they label (1)—(3), and since they agree that (3) is a logical consequence of the first two, I need consider only their responses to (1) and (2).

(1) Against the claim that an action that takes (or goes on for) a stretch is occurring at any moment in that stretch, Weil and Thalberg suggest that singing may go on all afternoon although there are pauses during which vocalizing stops. Why do they switch from 'singing' to 'vocalizing' here? Consider how their point would go if we stayed with 'singing': someone's singing may go on all afternoon, although there are pauses during which his singing is not occurring. Really? Well, if we shift our standards of non-interruptedness mid-sentence, then we can say this, I suppose. But my premiss required to be read by reference to a single standard; any standard will do provided it is adhered to. Perhaps we can see now why Weil and Thalberg switched verbs: presumably they hoped to exploit the discrepant standards of non-interruptedness needed for the applicability of different verbs. Of course to exploit the possibility of such discrepancies in order to try to make out that there can be an action which goes on for a stretch *and* which does not occur for moments during that stretch, one needs to assume identities, e.g. between his singing, which is said to go on all the time, and his vocalizing, which is said not to occur at certain moments. I do not know what Weil and Thalberg think about the identity of his singing with his vocalizing; and I lack clear intuitions about vocalizing myself. But I do think that there can be identities between actions picked out using verbs that differ from one another in respect of the standard of non-interruptedness needed if they are to be truly said to apply for a stretch. (Rather as I think that there can be identities between some buildings and some skyscrapers, even if a different height is needed to have a tall building from that needed to have a tall skyscraper.[2])

(2) In saying that any moment at which *a*'s action is occurring is a moment when *a* is doing something, I intended the use of 'do something' instances of which are action verbs that can be used to name things that are done. (Consider: 'What did he do?—Sing: something he did was sing.') Granted that there is such a use, and that it serves simply to schematize, it is hard to see how my premiss could be other than trivially true. Weil and Thalberg's example is out of step with that use: 'her laundry' is not a verb.[3] Of course 'do her laundry' is a verb. But then,

---

[2] I do not say that the cases of 'uninterrupted' and 'tall' are perfectly parallel. One important difference is that standards for non-interruptedness not only differ from verb to verb, but may shift even when a single verb is used. See David Lewis 'Scorekeeping in a Conversation Game', *Journal of Philosophical Logic* 1979, pp. 339–59 for an illuminating account of the phenomenon by which standards of precision may shift in the course of a conversation. One would need to adduce that phenomenon to give a complete account of the behaviour of verbs.

[3] These and other senses of 'do something' are distinguished by Judith Jarvis Thomson in a work that Weil and Thalberg cite: *Acts and Other Events* (Cornell University Press, 1977).

405 An 13

with that as the relevant instance, my claim stands. For *if* it is literally true that she is doing her laundry as she listens to the machine, then as she listens to the machine, she is doing something: something she is doing is doing her laundry.

IV

Let me end with a terminological point, affecting even Weil and Thalberg's title, whose general recognition would seem to be nothing less than a *sine qua non* of progress in the philosophy of action. If one thing someone did were a basic action, then obviously it could not be the same as something he did if the latter were a non-basic action. No one would make such a claim. This shows that 'action' is not used univocally in the phrase 'basic *action*' and in claims about *action* identities. The subjects of the disputed identity claims are people's doings of things (which are particulars), whereas the candidates for more or less basic items are the things that people do (which are universals). With that in mind, one is in a position to see that a Unifier holds merely that someone's doing some basic thing can be (the same as) his doing some non-basic thing. Then one will not wish to land him with the view that actions are 'misleadingly' described when described as the doings of non-basic things (p. 12). As if it were misleading to say that I have written a reply to Weil and Thalberg's paper.

Corpus Christi College, Oxford © JENNIFER HORNSBY 1981

P 1849

ON AN OBJECTION TO THE SYNONYMY PRINCIPLE OF PROPERTY IDENTITY

By MICHAEL TYE

UNTIL recently many philosophers accepted the following principle of property identity (or something very like it):

(I) Property *P* is identical with property *Q* iff given any predicates *F* and *G*, if *P* is expressible by *F* and *Q* is expressible by *G* then *F* is synonymous with *G*.

This principle, (I), has been widely criticized in the last twenty years. Sometimes it has been argued that the concept of synonymy is hopelessly vague and hence that (I) provides no clear identity conditions for properties.[1] More frequently perhaps, it has been charged that there are straightforward counterexamples to (I).[2]

In what follows, my concern is with some of these alleged counterexamples. More specifically, I wish to defend (I) against the claim that it is refuted by various contingent property identities. I offer this defence for two reasons. First, I think it is safe to say that if the recent literature on property identity demonstrates anything at all, it demonstrates how difficult it is to formulate a satisfactory alternative to (I). Secondly, it seems to me that (I) is a plausible principle of property identity, one which merits further consideration at least by those philosophers who are disinclined to reject the notion of synonymy as unintelligible.

I

It is often alleged that statements such as

(1) Red is the colour of ripe tomatoes,
(2) Blue is the colour of the sky,

and

(3) The property designated by the English word 'redness' is the same as the property designated by the German word 'Rot'

furnish contingent counterexamples to (I). Indeed, some philosophers seem to think that it is *obvious* that cases like these undercut (I). I shall begin my defence by arguing that neither claim is correct.

---

[1] See, e.g., W. V. O. Quine, *Word and Object* (M.I.T. Press, 1960), pp. 200–210.

[2] See, e.g., Richard B. Brandt, 'Doubts about the Identity Theory', in *Dimensions of Mind*, ed. by S. Hook (New York University Press, 1960), p. 66; Max Deutscher, 'Mental and Physical Properties', in *The Identity Theory of Mind*, ed. by C. F. Presley (University of Queensland Press, 1967), pp. 74–75; Jaegwon Kim, 'On the Psycho-Physical Identity Theory', *American Philosophical Quarterly*, 3 (1966), pp. 232–33; also 'Events as Property Exemplifications', in *Action Theory*, ed. by M. Brand and D. Walton (Reidel, Dordrecht, 1976), p. 168.

Consider, for the moment, only statement (1). The identity in (1) holds between the property designated by the name 'Red' and the property designated by the definite description 'the colour of ripe tomatoes'. Since the latter term is, in Kripke's phrase, a 'nonrigid designator' (one which picks out different entities in different possible worlds[3]) it is certainly true that (1) describes a contingent property identity. Note, however, that (I) makes no mention of *singular terms* for properties, that is, terms like those which flank the word 'is' in (1). It cannot be *immediately* obvious, then, that (1) refutes (I). Indeed, unless some link can be forged between the properties designated by the singular terms in (1) and the properties expressible by certain as yet unspecified predicates, the claim that (1) is a counterexample to (I) has no foundation at all, obvious or otherwise.

Presumably the relevant predicates here are 'is red' and 'has (or is) the colour of ripe tomatoes'. An argument against (I) can now be constructed as follows:

(4) 'Red' designates the property which 'is red' expresses.
(5) 'The colour of ripe tomatoes' designates the property which 'has the colour of ripe tomatoes' expresses.
(6) 'Is red' is not synonymous with 'has the colour of ripe tomatoes'.

(4) and (5), in conjunction with (1), entail

(7) 'Is red' expresses the same property as 'has the colour of ripe tomatoes'.

And (7), in conjunction with (6), yields the conclusion that (I) is mistaken.

This argument is not compelling, however. While premises (1), (4), and (6) should not be disputed, the defender of (I) can reasonably challenge premise (5). Let me explain.

The definite description 'the colour of ripe tomatoes' designates the property which uniquely satisfies it, that is, redness (or being red). The predicate 'has the colour of ripe tomatoes' expresses the property of having the colour of ripe tomatoes just as the predicate 'is red' expresses the property of being red. It follows, then, that (5) is not true unless

(8) Being red is having the colour of ripe tomatoes

is true. But (8) is not, I think, the same as (1) although it is very easy to confuse the two. Indeed, in my view, (8) is false. Why? Two major reasons.

To begin with, it seems to me that singular terms of the form 'being the *F*' or 'having the *F*' are rigid designators. I shall take up the consequences of this claim for (8) shortly. First, some preliminary comments are necessary. In those cases where the definite descriptions embedded in the singular terms are rigid designators, my claim is uncontroversial. But, in the other cases, where the embedded definite descriptions are nonrigid, there may appear to be a simple mistake. After all, it may be objected, if a description of the form 'the *F*' is nonrigid, then a singular term of the form 'being the *F*' or 'having the *F*' which contains the description must be nonrigid too. My reply here is that this inference is invalid, as the following discussion of the example 'being the inventor of bifocals' shows: Benjamin Franklin is the actual inventor of bifocals. Suppose that another man—George Washington, say— is the inventor of bifocals in some possible world much like the actual one. Then intuitively there is some common property which is attributed to these men at their respective worlds by users of the predicate 'is the inventor of bifocals' in those worlds, namely being the inventor of bifocals. Hence, the singular term 'being the inventor of bifocals' designates the same property in both worlds. Hence, generalizing, we arrive at the conclusion that 'being the inventor of bifocals' designates the same property in all worlds despite the fact that the embedded description 'the inventor of bifocals' is nonrigid.

---

[3] For more on nonrigid (and rigid) designators, see Saul Kripke, 'Naming and Necessity,' in *Semantics of Natural Language*, ed. by D. Davidson and G. Harman (Reidel, Dordrecht, 1972), pp. 269–90.

This sort of reasoning may be applied to other designators of the form 'being the *F*' or 'having the *F*'. In particular, it may be applied to 'having the colour of ripe tomatoes'. To see this, consider a red object, α, in the actual world and a green object, β, in some possible world in which green is the colour of ripe tomatoes. Intuitively, there is a property which is shared by α and β in virtue of which the *same* predicate ('has the colour of ripe tomatoes') may be correctly applied to each object at the appropriate world. This property is the one designated by 'having the colour of ripe tomatoes'. Generalizing, it seems plausible to suppose that 'having the colour of ripe tomatoes' designates the same property in all worlds—a property which is common to all and only red objects in the actual world, to all and only green objects in the given possible world, and so on. Hence, on this view, 'having the colour of ripe tomatoes' is a rigid designator.

Now clearly 'being red' is also a rigid designator. Hence if (8) is true, then it is necessarily true. Reflection on the possible object β shows that (8) is *not* necessarily true. Hence, we are led to the conclusion that (8) is not true at all.

My second reason for rejecting (8) can be stated very briefly. On the standard account, properties are attributes, that is, entities which may be introduced via predicates to *characterize* the objects of reference.[4] According to my intuitions, however, the property of having the colour of ripe tomatoes is *not* the attribute with which a person characterizes an object when, in the ordinary course of events, he says that it is red. Hence, as above, it seems to me that (8) is false.

---

[4] This is not to say, of course, that properties cannot also be referred to by singular terms.

The conclusion I draw, then, is as follows. The contingent property identity in (1) undermines (I) if statements (4), (5), and (6) are true. But (5) is true only if (8) is true, and a defender of (I) may reasonably maintain that (8) is not true. Since there appears to be no way of using (1) against (I) without introducing (5) (or some other parallel and hence equally vulnerable assumption), it seems to me that (1) does not compromise (I). The same conclusion holds, I suggest, for all contingent property identity statements with the grammatical form of (1) and (2), be they drawn from everyday or scientific contexts, and also, by similar reasoning, for all such statements with the grammatical form of (3). I might add relatedly that the position I have adopted with respect to (8) can be generalized too. That is to say, if (8) is false, as I have argued, then it seems to me that all statements which allegedly represent contingent property identities, and which have the grammatical form of (8), do not represent genuine identities at all. In consequence, they pose no threat to (I).

Are there *any* contingent property identities which refute (I)? I think not. It seems to me that putative contingent counterexamples to (I) which do not exactly fit the pattern of the identity statements I have discussed will still be open to objections like those presented above (and, in some cases, to other objections too). By way of illustration, consider the following scientific identity:

(9) Being at 0°C is being at 273°A.[5]

According to some philosophers, this is a contingent identity which derives from the ways the Centigrade and Absolute temperature scales happen to have been set. (9), then, supposedly is a counterexample to (I), since the predicates 'is at 0°C' and 'is at 273°A' are not synonymous.

How the defender of (I) replies to this case should depend upon how the singular terms '0°C' and '273°A' are classified. Suppose, initially, that these terms are rigid connotationless names; or, in other words, that '0°C' and '273°A' are taken to designate but not to describe the same temperature in all possible worlds. Then, it follows that (9) is really a necessary truth and not a contingent one. Moreover, on this reading, the predicates 'is at 0°C' and 'is at 273°A' *are* synonymous.[6] So (9) is not a counterexample to (I), contingent *or* necessary.

---

[5] Deutscher presents an identity like this one in his 'Mental and Physical Properties', *op. cit.*, pp. 74–5.

[6] It may be objected that if these predicates are taken to be synonymous, then (9) cannot represent a scientific discovery. For (9), on this account, is synonymous with the trivial statement 'Being at 0°C is being at 0°C'. This problem is not one which the defender of (I) is required to address, since it challenges the viability of the assumption that '0°C' and '273°A' are rigid connotationless names rather than the plausibility of (I) itself.

Suppose, however, that 'o°C' and '273°A' are viewed as covert definite descriptions, say, as 'the freezing point of pure water at normal pressure' and 'the temperature which is 273 units (on the Centigrade scale) above the temperature at which molecular motion ceases'. Then, it seems to me, the defender of (I) can argue that (9) is false. For the property now designated by 'being at 273°A', that is, the property of being at the temperature which is 273 units . . . is clearly *not* what a speaker normally attributes to an object when he characterizes it as being at o°C. The attributed entity in this case, given the descriptive view of 'o°C', is being at the freezing point of pure water at normal pressure. And intuitively an ordinary speaker who attributes *this* property to an object is not thereby attributing to it a complex molecular property which likely he is entirely unfamiliar with.

This point parallels one of the reasons I gave earlier for denying that (8) is true.[7] And just as the earlier denial of (8) did not require the rejection of (1), so this denial of (9) does not require the rejection of

(10) o°C is 273°A.

Hence, the claim that (9) is false (on the view that 'o°C' and '273°A' are covert descriptions) is not obviously counterintuitive or perplexing.

That concludes my defence of (I). Those who believe that there are contingent property identities which refute (I) now face the task of producing some compelling examples which are not susceptible to the sorts of objections I have offered.[8]

Northern Illinois University © MICHAEL TYE 1981

---

[7] My other objection to (8) can also be extended to (9) without much difficulty, if 'o°C' and '273°A' are taken to be covert definite descriptions (and (9) is proposed as a contingent truth).

[8] This paper was accepted for publication before the appearance of Frank Jackson, 'A Note on Physicalism and Heat', *Australasian Journal of Philosophy*, 58 (1980), pp. 26–34. My first argument against (8), although developed independently of Jackson's work, is very similar to the one he presents on p. 29, *ibid.*

A NOTE ON THE SUBSTITUTIVITY OF NOTES

By G. BAR-ELLI

LEIBNIZ's famous principle of Identity, in one of its versions, says that two terms that denote the same object are substitutable for each other in any context *salva veritate*. Its compelling intuitive force notwithstanding, it is notorious that there are cases which seem to be counter-examples to the principle: belief contexts can illustrate the point nicely, for it can plainly be the case, for example, that a equals b, and that John believes that a is F, and yet John may not believe that b is F. It is worth noticing, however, that in all such cases there seem to be particular elements which "spoil" the extensionality of the context; elements which are "responsible" for the seeming failure of the principle of substitutivity in that context: be it a propositional verb (like 'believes', 'knows', etc.), or some modal operator (like 'necessarily'), or some causal verb, etc. This seems to me important because in the case I bring below no such element is discernible.

By and large, philosophers have tried to accommodate those contexts to the basic insight of Leibniz's law by claiming that where the principle seems to fail, this is because it is misapplied: be it by confusing direct and indirect reference (Frege), or by confusing positions that are "purely referential" with those that are not (Quine).

Consider now the following case: in classical, tonal music 'A-sharp' and 'B-flat' denote the same thing, the same musical sound. This is not strictly so in music for strings, or in vocal music, but it is so in music written for most keyboard instruments, and for our purposes this is all we need be concerned about. But substitutivity evidently fails with regard to these terms: 'A-sharp is a leading tone in a B scale' is a truth which turns into falsehood when we substitute 'B-flat' for 'A-sharp'. This is not an exceptional case: substitutivity fails in almost every normal use of these terms. To give one more example, a G minor chord is composed of G, B-flat, and D, and not of G, A-sharp, and D.

How are we to explain this apparent failure of substitutivity? The contexts in question are perfectly straightforward, "direct" descriptive contexts, which do not include any "intensional", or "oblique" element, and which do not seem to call for any reference-shift as required, e.g. by the Fregean approach to belief and other opaque contexts. It also seems clear that in these contexts the occurrences of the terms involved are perfectly referential, so that any conclusion of non-referentiality in the Quinean way of handling such contexts would be totally *ad hoc* and unilluminating. The answer to our question cannot, therefore, take these lines.

I believe that this trifling puzzle has far-reaching consequences for some aspects of musical notation, for the nature of what may be called "musical ontology"—the referents of musical terms—and in fact for the nature of music in general. I shall try to spell out a few of these consequences in rough outline, but first some preliminary remarks seem to be in order.

(1) Musical notation can be looked upon in various different ways. The one I assume here may be epitomized by saying that *the notes denote*:[1] they are designative terms; they denote some objects (the appropriate tones) and in using them in statements about music we strictly refer to these objects. In talking here of muscial statements I do not mean played or heard phrases in music: I mean rather ordinary linguistic statements about music, e.g. that a G minor chord is composed of G, B-flat and D. There is, obviously, a difference between English expressions like 'B-flat' and the corresponding notes;[2] but this is, I believe, of little moment to my argument: in talking about expressions like 'B-flat' one gains generality, which strict musical notation lacks. A statement like the above one about the G minor chord cannot be written using notes instead of 'B-flat'. But whatever I say here using the English expressions can be applied, with obvious changes, to the notes themselves. As examples, one may take here particular instances of the above general statements, e.g. 'Beethoven's "Pathétique" sonata begins with a C-minor chord, which consists of . . . ' does not permit substituting, say, the note of D-sharp for that of E-flat. Hence, at some points in the sequel I blur the distinction and apply my conclusions about expressions like 'B-flat' to the notes themselves.

(2) These musical statements, which contain musical terms, are strictly true or false: that a G minor chord contains B-flat is true, and that it contains A-sharp is false, and truth and falsehood are not used here metaphorically, but strictly. Truth and falsity are not conventional in these cases more than in any other ordinary statements of empirical facts. Musical notation and terminology are conventional in the same way as any system of notation and any terminology are: we could have called the tones differently, or we could have described the musical phenomena in a totally different way. But this is true in any domain, and of any language. The point about truth here is that given our musical terminology, and given what these terms mean, and the way they are used, truth and falsity are as strict here as anywhere. That a G minor

---

[1] I use 'notes' the American way to refer to *written signs* in musical notation. The word 'tone' I reserve for what is denoted by these notes. It would be true to point out that tones are not physical events but, if at all, sets of such events; also, that notes are, in a way, indexical because they do not imply 'absolute pitch', etc. I ignore all these niceties because they do not seem to be relevant to my argument.

[2] Here I am indebted to the Editor of ANALYSIS.

chord contains B-flat and not A-sharp is not more "conventional" than is the fact that two plus two are four, not five. In both cases we could have talked otherwise, and used our terms differently, but given the way we actually use them and what they mean, that is how things stand.

(3) I should perhaps add here that A-sharp is not less common, or more bizarre, a tone than B-flat. They are equally normal and common, and they can both appear in the same work in the same phrase. The fact that A-sharp is not contained in a B-minor chord is not due to any bizarreness of use, as e.g. it may be bizarre to refer to a colour in a painting by its chemical formula, or as it would be wrong for me to address the Queen of England by her first name. This "wrong" is really conventional in a straightforward way, but our musical case is totally different when I say that it is wrong that a B chord contains A-sharp, the "wrong" means simply and strictly 'not true'.

I should also add that the examples I have brought are perfectly regular and typical cases, and endless numbers of examples like them can be produced at will. Moreover, that is so not only with regard to single tones but with regard to other musical objects as well: the same chord (in the sense of the same physical event) can be denoted by two different names and written by different notes which will not, in general, be interchangeable in the same musical statement. A similar point can be made about rhythmical patterns.

Granted these remarks, we can now come back to our initial puzzle. Throughout the discussion we have assumed that the referents of 'A sharp' and of 'B-flat' (and of the corresponding notes) are the same, that they refer to the same tone. On the basis of this assumption I have tried to show that there is no easy way of dealing with our puzzle: Leibniz's law is compelling and there does not seem to be any reason to think that it is inapplicable to the contexts we are dealing with. It is time then to examine the initial assumption about the identity of A-sharp and B-flat more closely.

There seem to be strong reasons in favour of this identity. These expressions and the corresponding notes are, as we observed, denoting terms; their referents—the elements of musical ontology for which they stand—are, supposedly, physical events, or sets of such events (be they acoustic or neurophysiological, or what have you), and producing the physical event that corresponds to A-sharp would have exactly the same result as that of producing B-flat. Hence, it seems that, other differences notwithstanding, as far as their *referents* are concerned 'A-sharp' and 'B-flat' denote exactly the same thing.

However, the inevitable conclusion of our discussion, it seems to me, is that in spite of its seeming plausibility this statement of identity cannot be correct, and A-sharp and B-flat are not, after all, the same tones; and hence, tones (which are the referents of notes) are not identical with physical events. The physical events which correspond to A-sharp and to B-flat are the same, but the tones are not. The tones are not merely physical sounds; they, with the rest of musical ontology, are 'intensional' in a way that makes A-sharp different from B-flat, in spite of their identical corresponding physical events. This much is established I believe, by our previous considerations. But these considerations are certainly not sufficient to *explain* this difference, and to formulate in what exactly it consists. I cannot here venture to provide such an explanation, but I should like to point in what seems to me the right direction.

Musical terms, and musical terminology, one may suggest, are *theoretical*, wherein lies their main difference from ordinary proper names, and wherein one should search for the desired explanations of the above facts. But even if there is some truth in this answer there is much that it misses. For, although it is true that musical terms are incorporated and 'defined' in a rather articulated theory, so are, for example, the numerals, and a host of other terms, which do not display the same kind of opacity. Moreover, it is clear that if we had terms designating the physical events corresponding to the musical tones, they would surely not be less theoretical than the musical terms themselves. So, being theoretical cannot explain their difference in behaviour.

It is rather the *kind* of theory we have to deal with which is important here. Musical terms, I would suggest, are 'functional' in a particular way in which other terms are not. This does not only mean that musical terms are incorporated and, perhaps, defined, in an elaborate theory; it rather means that the theory assigns them a particular role in a particular context: the same sound is associated with the different tones, according to its functions in different contexts.

This formulation is, however, too general to characterize musical talk in a sufficiently narrow way. For, it seems that other uses of terms, like mathematical use of the numerals, or even everyday use of names, may be 'functional' in the same way. A star, for instance, may be assigned two different names according to different functions it has in different contexts. This in itself would not prevent us, however, from claiming strict identity between the referents of these terms, substituting one for the other, etc. So the kind of functionality which is thus constitutive of music is different.

What is needed, therefore, is a more refined description of the kind of functionality characteristic of musical terms, which will explain their particular "intensionality". I cannot, of course, give such a description here, nor is it needed: we all know, more or less what it could be like. Generally, it consists of the basic principles of Harmony and Counterpoint, familiar to any student of music. More specifically, when applied to particular musical contexts, it consists of the minute details of musical analysis, which are also known to, and practised by, musicians with a high degree of agreement.

One may object here to my emphasizing the *functional* character of musical terms and suggest alternative, more modest, explanations of the identity of tones: cannot a tone be identified with a certain *type* of which the various physical sounds are tokens, where these types are specified in a way that secures the difference between A-sharp and B-flat? Or, alternatively, cannot a sound be identified with a tone by its place in a certain order?

I cannot discuss these and other suggestions in detail within the scope of this article. I should, however, indicate briefly where I see the propriety of my functional terminology. Take, for instance, the second of the above suggestions. What order is in question here? When one tries to spell it out in detail one realizes that the set of sequences that have to be taken into account here is indefinite; the only way I know of specifying it, even roughly, is by the functional theories of Music alluded to above. Moreover, even if it could be specified in some other way, it would then define the difference between A-sharp and B-flat in a way that might be extensionally adequate but that would lack any *explanatory* power, which the functional theories and the functional characterization of these tones have. These do not only define the difference between A-sharp and B-flat, but also explain the point of this distinction; they also explain why a particular tone occurs at a particular place instead of its enharmonic substituent.

The standard available theories of Harmony, Counterpoint, Rhythm, etc. may not be complete: they may all need refinements and revisions, and they may undergo fundamental changes, and "revolutions". But in this they do not differ from any other theory, in any other domain. Some people may object here that what I call musical theories are, in fact, nothing but sets of rules, or 'manuals for correct writing', which are entirely normative in character, and hence that the descriptive role I ascribe to them in characterizing the particular functionality of musical tones is misconceived. This is a large question, which cannot be discussed here in detail. I should only say, rather dogmatically, that such a view seems to me highly inadequate: the principles of those musical theories are right or wrong; they are more or less fruitful in their consequences, like any other principles of 'established' theories.

In the above I have assumed that the notes denote and that they are not, say, merely performative instructions. But now, having made my point, we can see why they should be construed that way. For, granted a 'realistic' attitude towards musical theories, and viewing them as descriptive and explanatory theories of (part of) the musical phenomena (within the relatively well-defined region of classical tonal music) we see that *there is* a difference between A-sharp and B-flat, etc. and that the notational distinction between them is fully justified and substantial. But if the notes were nothing but performative instructions, this distinction would be arbitrary and pointless. This can also show us that notational systems whose performative utility may be equivalent to the standard one, but which make no distinction between A-sharp and B-flat, etc. are quite deficient in comparison to the standard notation, for they give an inadequate description of the musical phenomena. This is not to deny that other kinds of notation may fit better other kinds of music, for which the distinction between A-sharp and B-flat, for instance, is not vital.

The Hebrew University of Jerusalem © G. BAR-ELLI 1980

AN ANTI-ESSENTIALIST NOTE ON SUBSTANCES

By GRAEME FORBES

AMONGST the variety of essentialist doctrines, we can distinguish between those which do, and those which do not, invoke the notion of a natural kind. One which does not is Kripke's view that organisms are essentially related to the entities from which they develop.[1] For examples of doctrines which do, one naturally thinks of Putnam's views about elements and compounds, for instance, his doctrine that necessarily, water is H_2O.[2] In general, according to Putnam, elements have their fundamental physical properties, such as atomic number, essentially, and compounds have their fundamental structure, their molecular structure, essentially. In this paper, I wish to raise some queries about how these doctrines are to be construed.

The thesis that necessarily, water is H_2O, is hardly univocal. Of course, if we treat the mass term 'water' as a name (of a substance), and 'H_2O' as a name also, in the manner of Kripke (op. cit. p. 323), the reading of the thesis as an identity sentence governed by a modal operator is forced upon us. But there is good reason, independent of views about the essence of substances, to treat mass terms as predicates;[3]

---

[1] See 'Naming and Necessity' in *Semantics of Natural Language* eds. Davidson and Harman (Reidel 1972).

[2] Expounded in many places. I shall refer to his 'Meaning, Reference and Stereotypes' in *Meaning and Translation* eds. Guenther and Guenther-Reutter (Duckworth 1978) pp. 61–81.

[3] See Tyler Burge's 'Truth and Mass Terms', *Journal of Philosophy* vol. 69, especially pp. 271–6.

and as soon as we do this, we can discern interestingly different ways of formalizing 'Necessarily, water is H_2O'. Suppose we also treat 'H_2O' as a predicate (this is just for expository convenience): then we can have both

(1) $\boxed{\scriptstyle a}(\forall x)\ (\mathrm{Water}(x) \rightarrow H_2O(x))$

and

(2) $(\forall x)\ (\mathrm{Water}(x) \rightarrow \Box(\mathrm{Exists}(x) \rightarrow H_2O(x))$.

(1) is perhaps the best representation of the thought which prompts the idea that 'Water is H_2O' is an identity sentence, the truth of whose necessitation is delivered by the necessity of identity. (2), on the other hand, is a claim about individual quantities of water, such as the water in the glass now on my desk: it implies that that quantity of stuff cannot exist without having the chemical structure H_2O.

Let us begin with (2). The water in my glass has certain contingent properties; for example, if I had stood it in the fridge before bringing it to my desk, it would be colder than it actually is. But could it have had a molecular structure other than H_2O? This is a question to which Putnam's best known arguments are not addressed. Recall the famous Twin Earth (TE) thought experiment (op. cit. pp 62–5): what is at issue there is whether an arbitrary quantity of substance in a world w would be properly *classified* (in English, with respect to w) as water, if it did not have the structure H_2O in w. What is not at issue is a claim about what conditions have to be satisfied if arbitrary quantities of substance in a pair of worlds are to be correctly *identified* across worlds as the very same stuff. Nevertheless, it is plausible that (2) is true. Otherwise, we may suppose that there is some world w in which some entity x exists which exists in the actual world and is water, and therefore H_2O, in the actual world, but which has a different molecular structure in w. This claim is of a kind with other, unintuitive, anti-essentialist ideas. Suppose a certain oak tree *a* actually originated from an acorn *b*; imagine that we are in possession of sufficiently sophisticated genetic engineering to have been able to plant *b* and ensure that a small vegetable developed from it. Does this show that a certain oak tree could have been, say, a cabbage? If it does, then we must have the conception of a subject which exists in two worlds, and is an oak tree in one and a cabbage in the other. It is difficult not to believe that we have no such conception; and that such a conception is unattractively close to the scholastic notion of a substance as the bare substratum in which an individual's properties inhere.[4] (2) is therefore one of a number of plausible essentialist theses which impute essential properties to *individuals* on the basis of the kind to which they belong. This is one way in which essentialism may employ the notion of a natural kind.

---

[4] This view is also expressed by Wiggins in his 'Ayer on Monism, Pluralism and Essence' in *Perception and Identity: Essays Presented to A. J. Ayer* ed. G. Macdonald (Macmillan 1979).

The other form is exhibited in (1), to the establishing of which Putnam's TE thought experiment is directed. I wish to argue that the conclusions which may be legitimately drawn from the experiment fail to show that (1) is true, and that there is independent motivation for scepticism about (1). Putnam asks us to imagine circumstances in which English-speaking explorers leave the planet Earth and arrive at a planet, Twin Earth, which is an epistemic counterpart of Earth.[5] On Twin Earth, there is a substance quantities of which have all the stereotypical features of quantities of what our explorers call 'water' on Earth; but on Twin Earth, this epistemic counterpart of water is XYZ. Putnam claims, and I agree, that the explorers should say that they have discovered some substance which isn't water, but which is very like it. This shows, he says, that in the mouths of the explorers, and so in English, 'water' has an 'unnoticed indexical component'. So far, I am still in agreement; but our paths diverge when it comes to specifying what the example suggests the indexical component to be. Putnam draws two morals. The first is that water is stuff which 'bears a certain similarity relation to the water around here' (op. cit. p. 71), and this seems right. The indexical 'here' is a *spatial* indexical, picking out a certain region of space-time; putting equivalent words into the mouths of our explorers on Twin Earth, we could perhaps use the description 'the place where we come from'. The other moral which Putnam draws, however, does not follow from the agreed claim about what the explorers should say: the purported moral is that the indexical to be associated with 'water' is a *modal* indexical—water is 'whatever bears a certain equivalence relation . . . to the piece of liquid referred to as "this" in the actual world', where the relevant use of 'this' picks out a paradigm sample of water (op. cit. p. 68; in speaking of a modal indexical, I am merely conforming with Putnam's terminology—I do not hold that the semantic properties of backwards-looking operators in modal languages are to be explained in terms of context-dependence). So far as I can see, Putnam gives no argument for deriving this latter moral; he just seems to assume that if the first can be drawn, so can the second:

> Suppose, now, that I discover the microstructure of water . . . that water is H_2O. At this point, I will be able to say that the stuff on Twin Earth that I earlier mistook for water is not really water. *In the same way*, if you describe, not another planet in the actual universe, but another possible universe in which there is stuff with the chemical formula XYZ which passes the 'operational test' for water, we shall have to say that that stuff isn't water but merely XYZ. (Op. cit. p. 70, my emphasis.)

---

[5] For the notion of an epistemic counterpart, see Colin McGinn's 'Anomalous Monism and Kripke's Cartesian Intuitions', ANALYSIS 37.2, January 1977, pp. 78–80.

But the TE example, by itself, does not force us to say this at all. There are two distinct objections which can be made here. One is that the explicit moral which Putnam draws, that in any world 'water' applies to something iff that something stands in the *same liquid* relation to a certain quantity of water in the actual world (op. cit. p. 68), is a moral too weak to imply that water could not have been anything but H_2O: we need the additional premiss that *what it is* to stand in the *same liquid* relation across worlds is to have the same molecular structure.[6] But this is not my objection, since I anticipate the reply that this shows only that the correct moral to draw is, e.g. that in any world 'water' applies to something iff that something has the same fundamental physical properties as a certain piece of water in the actual world. My objection, rather, is that the TE example does not support *any* principle which constrains the crossworld application of 'water'. In that example, explorers go from planet to planet, not possible world to possible world; so the most we can derive from a decision about what it is correct for them to say is a constraint on the cross-physical-space application of 'water', roughly to the effect that throughout the physical space of each individual possible world, the application of a substance term is answerable to the fundamental physical properties of paradigm instances of the substance to be found at the place where use of the term originated. The same substance cannot have different-structured instances at different places. So we can assert only

$$(3)\quad \Box((\exists x)\,(\mathrm{Water}(x)\ \&\ \mathrm{H_2O}(x)) \rightarrow (\forall x)\,(\mathrm{Water}(x) \rightarrow \mathrm{H_2O}(x)))$$

which is consistent with the existence of a possible world where every sample of water is XYZ.

Can we improve the case against (1) beyond this argument against an argument? Perhaps so. One advantage of formulating essentialist theses about substances as the necessitations of identity sentences is that the necessity of identity guarantees the truth of the theses.[7] But if 'water' is a rigid designator, what does it designate? Not the set, or mereological sum, of all quantities of actual water, since, presumably, there could have been more or less water than there actually is. The only plausible candidate seems to be *waterhood*, the abstract property of being water, but this choice of referent is vulnerable to the objection, due to Donnellan, that on such an interpretation, the proposition that water is H_2O is not one which science has discovered to be true; science has only discovered that the things which actually have the property of being water are the same as the things which actually have the property of being H_2O.

---

[6] There is an illuminating discussion of crossworld relations in Nathan Salmon's forthcoming book *Reference and Essentialism*, where Salmon makes this objection to Putnam's argument.

[7] In 'Identity and Necessity' (in *Identity and Individuation*, ed. Munitz, New York, 1971, pp. 135–64) Kripke treats 'Heat is the motion of molecules' as an identity sentence, says that its necessitation is established by the necessity of identity, and gives Putnam-like thought experiments to convince us that its necessitation is true. In a commentary on that paper, I have tried to apply considerations, similar to those presented here, against Kripke's arguments.

Nevertheless, we can imagine a position which says that if we take into account more general features of transworld identity, then just the actual coextensiveness of the properties is sufficient to establish (1) if the predicates in (1) express those properties. The position is this. Suppose that through a period of time we uninterruptedly observe a scene which gives every appearance of the continuous existence of some material object, say a lectern, undergoing no perceptible change. According to common sense, such sensory information is at least *prima facie* evidence that the lectern we observed at the beginning of the period was the same lectern as the one we were observing at the end. It would be bizarre to insist that we do not have reasonable grounds for such a judgement until we have gathered evidence about the states of *causally unrelated* lecterns at the end of the period, elsewhere in the physical universe. The common sense view is that identity through time is intrinsic, and the bizarre view, that it is extrinsic. Let us grant the common sense view and grant the essentialist that identity through worlds is like identity through time in this respect. Thus someone who thinks that the identity of a neutrally described individual in a world can turn on the properties of other causally unrelated individuals in that world, say because he fixes transworld 'identities' by some notion of *comparative* similarity, offends the intrinsicness of transworld identity; whereas someone who uses the identities of the entities from which the neutrally described individual originates in the world, if it is an organism, or from which it is fashioned or constructed in the world, if it is an ornament or artefact, does not. A defender of (1) may now claim that if the transworld identity conditions of an abstract object such as the property of being water are not extrinsically determinable, there must be something which fixes the identity of that property in a world just as the sperm and the egg from which a human originates in a world fix the identity of that human. And what is more natural than the suggestion that the identity of a substance property picked out with respect to a world by some of its characteristics there (e.g. that things which have it are thirst-quenching) is fixed by the identities of the fundamental physical properties of the things which have the neutrally described property?

According to such a view, the distinction with which this paper began between types of essentialist doctrines does not mark an important difference: both types rely for their support on the same feature of transworld identity. Perhaps my conclusion should just be that we need to see the suggested parallel drawn out in fine detail, and until that is done, we have not been given any compelling reason to believe (1). However, when we are concerned with abstract objects such as substance properties, there seem to me to be grounds for scepticism that such reasons will be forthcoming from an application of the intrinsicness constraint, to do with the question of whether the concept of transworld identity for such abstract objects is sufficiently well articulated to entail that someone who denies (1) is committed to *falsehoods*. It may not be definitely *correct* to say that if matter had been non-atomic, so that there were no molecules anyway, but some substance had exhibited all the characteristics of water compatible with that, then that substance would have been water. But it may not be incorrect either; there may simply be no fact of the matter.[8]

New College, Oxford © GRAEME FORBES 1981

---

[8] I have benefited from conversations with Martin Davies and Christopher Peacocke on the topic of this paper, and from correspondence with Nathan Salmon and the editor of ANALYSIS. Discussion with David Kaplan has been especially helpful.

LEWIS AND THE PROBLEM OF CAUSAL SUFFICIENCY

By DAVID-HILLEL RUBEN

IN his article 'Causation',[1] David Lewis believes, or anyway assumes for the purpose of discussion, that a cause is, in the circumstances, both necessary and sufficient for its effect, the necessity only being qualified to take into account cases of pre-emption. For simplicity of exposition, I want to avoid those later qualifications and the causal chains in which they would involve us. This simplification is quite harmless for the purposes of this paper, since it is the sufficiency thesis alone which shall concern me. There has been some discussion of the counterfactual analysis of causation and the necessity of a cause for its effect,[2] but none, that I am aware of, about the counterfactual analysis of causation and the sufficiency of a cause for its effect.

Lewis nowhere uses the terminology of a cause being either necessary or sufficient in the circumstances for its effect, but I think that he is committed to these views about causation nonetheless, and that it will be instructive to consider his counterfactual analysis of causation in these terms. It is perhaps easier to see that Lewis believes that a cause is necessary in the circumstances for its effect, since his analysis of causal dependence for two actual events, c and e, is that if c had not been, e never would have existed (c was necessary in the circumstances for e). But the sufficiency in the circumstances of a cause for its effect is also a part of Lewis' analysis, and rightly so, for nothing could be an adequate analysis of the common notion of causation that omitted this feature. If an analysis failed to catch sufficiency, it would fail to catch causation, and Lewis would, I think, agree that this was so.

Lewis says that he is 'concerned with the prior question of what it is to be one of the causes (unselectively speaking)' (p. 183). Thus, for Lewis, an event has more than one cause, and this might seem to count against my assertion that, for him, sufficiency is part of the analysis of causation. It does not, for the claim is only that, *in the circumstances*, which include the occurrence of the other causes, a cause is sufficient (and necessary) for its effect, and this can be true for each one of the multiple causes of an event. Lewis nowhere uses the explicit qualification 'in the circumstances', but the truth-conditions he offers for counterfactuals render that explicit qualification unnecessary. 'If c had occurred, e would have occurred' (when c and e did not actually occur) is non-vacuously true if some c-world in which e occurs is overall more similar to the actual world than any c-world in which e fails to occur. But the most similar c-world *is* a world in which all circumstances other than the occurrence of c and e (and whatever their occurrence entails) are the same as in the actual world, and this has the same effect as what, in an older terminology, would have been termed 'sufficient in the circumstances' (or, 'necessary in the circumstances'). In what follows, I write 'sufficient$_c$' and 'necessary$_c$' with a subscript 'c', to indicate that this qualification is being understood.

---

[1] David Lewis, 'Causation', *Journal of Philosophy*, 1973, pp. 556–67. Reprinted in Ernest Sosa, ed., *Causation and Conditionals*, Oxford, 1975, pp. 180–91. Subsequent quotations are from the Sosa volume.

[2] See for example: Louis Loeb, 'Causal Theories and Overdetermination', *Journal of Philosophy*, September 1974, pp. 525–44; Marshall Swain, 'A Counterfactual Analysis of Event Causation', *Philosophical Studies*, July 1978, pp. 1–19; William Goosens, 'Causal Chains and Counterfactuals', *Journal of Philosophy*, September 1979, pp. 489–95.

Some of the evidence for my contention that Lewis' causes are sufficient$_c$ for their effects is this. First, Lewis says for instance 'Suppose further that, given the laws and some of the actual circumstances, c could not have failed to cause e' (p. 189). I take 'could not have failed to cause' as 'was causally sufficient$_c$ for'. Second, before Lewis gives his formal, counterfactual account of causal dependence between particular events, he informally explains 'e depends causally on c' as 'whether e occurs or not depends on whether c occurs or not' (p. 187) and this thought seems to be made up of two parts: c's occurrence makes e occur and c's non-occurrence prevents e from occurring. I understand the first part as an informal rendering of causal sufficiency$_c$. Finally, when Lewis comes to compare possible worlds in terms of their comparative overall similarity, he counts what he calls 'minor violations of laws' and weighs them against differences in particular fact. Some of these minor violations of laws are cases in which effects occur in spite of the absence of their necessary$_c$ causes, but others are cases in which effects fail to occur in spite of the presence of their sufficient$_c$ causes.

The informal idea that 'whether e occurs or not depends on whether c occurs or not' (p. 187) is explained by Lewis more formally in terms of the truth of two counterfactuals:

> . . . $O(c) \;\Box\!\!\rightarrow O(e)$ and $\sim O(c) \;\Box\!\!\rightarrow \sim O(e)$. There are two cases. If c and e do not actually occur, then the second counterfactual is automatically true because its antecedent and consequent are true; so e depends causally on c iff the first counterfactual holds. That is, e would have occurred if c had occurred. But if c and e are actual events, then it is the first counterfactual that is automatically true. Then e depends causally on c iff, if c had not been, e never had existed.

This cannot amount to a more formal rendering of the informal 'whether e occurs or not depends on whether c occurs or not', or 'c is necessary$_c$ and sufficient$_c$ for e'. Let us take the second case that Lewis mentions, when c and e are actual events. When c and e are actual events, Lewis says that '$O(c) \;\Box\!\!\rightarrow O(e)$' is automatically true, and Marshall Swain, in a recent discussion of Lewis (op. cit. p. 4), points out that this is 'a consequence of Lewis' formal constraints on closeness of worlds', as indeed it is since no world can be more like the actual world than the actual world itself is. Thus, the causal dependence of e on c rests, in this case, on the truth of one counterfactual, 'if c had not occurred, e would not have occurred', and this counterfactual might capture the idea of the necessity$_c$ of a cause for its effect, but neither it, nor anything else in the analysis, captures the idea of the sufficiency$_c$ of a cause for its effect, the idea that whether e occurs depends on whether c occurs.[3]

It might be readily granted that, for the case in which c and e are actual events, '$\sim O(c) \;\Box\!\!\rightarrow\; \sim O(e)$' doesn't capture sufficiency$_c$—its job is to capture necessity$_c$—but it might not be as clear why '$O(c) \;\Box\!\!\rightarrow\; O(e)$' fails to capture the idea of sufficiency$_c$. The reason is this. Let a and b be any two actual events spatio-temporally related in whatever way one likes, but which are unrelated causally. Because they actually occur, this counterfactual is automatically true: 'If a had occurred, b would have occurred.' But since a and b are unrelated causally, the truth of this counterfactual cannot possibly be adequate for capturing the idea of causal sufficiency$_c$. The counterfactual '$O(c) \;\Box\!\!\rightarrow\; O(e)$', for the case in which the antecedent and consequent mention actual events, is for Lewis true just in case the corresponding material condition is, and if '$O(c) \supset O(e)$' could not explain the idea of causal sufficiency$_c$ in the case in which c and e occur, neither could '$O(c) \;\Box\!\!\rightarrow\; O(e)$'. I conclude then that Lewis' formal account in terms of the two counterfactuals cannot have got successfully at the heart of the informal idea that 'whether e occurs or not depends on whether c occurs or not', since in the case in which c and e do occur, that account cannot get at that part of the idea expressed by 'whether e occurs depends on whether c occurs'.

When c and e are actual events, the idea of causal necessity$_c$ can be explained as 'if c had not occurred, e would not have occurred'. We have just seen why causal sufficiency$_c$ cannot be explained as 'if c had occurred, e would have occurred'. A more natural way of explaining in counterfactual terms the idea of causal sufficiency$_c$ for actual events would be this: c is causally sufficient$_c$ for e iff if e had not occurred, c would not have occurred. I think that this would be the right way of explaining causal sufficiency$_c$ in counterfactual terms. This would mean that when c and e occur but are causally unrelated 'if c had occurred, e would have occurred' will be true, but 'if e had not occurred, c would not have occurred' will be false. This result would be acceptable to Lewis, who denies that a counterfactual and its contrapositive are equivalent (*Counterfactuals*, Oxford 1973, pp. 34–6).

---

[3] If we return for the moment to Lewis' first case, in which causation is defined for non-actual events, the analysis finds itself in a complementary difficulty over necessity rather than sufficiency. The idea of sufficiency in the circumstances might be captured by 'if c had occurred, e would have occurred', but since 'if c had not occurred, e would not have occurred' is automatically true when c and e do not occur, nothing in the analysis for this case captures the idea of the necessity in the circumstances of a cause for its effect. I have chosen to restrict myself in the text to the case in which c and e are actual events, in order to avoid becoming entangled in the question of whether the causal relation can hold between non-actual events (even assuming that there are such things). This point ought to increase our confidence in the belief that Lewis takes a cause to be sufficient$_c$ as well as necessary$_c$ for its effect. Lewis intends to provide an unambiguous analysis of 'c causes e' which applies both when c and e are actual and when c and e are non-actual. If we were to ascribe to Lewis the view that, when an actual c causes an actual e, it follows that c is necessary$_c$ for e but does not follow that c is sufficient$_c$ for e, then we also ought to ascribe to him the view that when a non-actual c causes a non-actual e, it follows that c is sufficient$_c$ for e but does not follow that c is necessary$_c$ for e. Clearly, this would be contrary to his intention of providing an unambiguous analysis that covered both cases.

However, it is still not open to Lewis to explain causal sufficiency$_c$ in this more natural way. Lewis is anxious to avoid making the causal relation symmetric, but is also anxious to avoid making it asymmetric simply by building into the analysis any explicit temporal requirement. In order to block the symmetry, he argues in several places[4] that when c and e occur and c causes e, 'if c had not occurred, then e would not have occurred' is true, but 'if e had not occurred, c would not have occurred' is false, on the grounds that the most similar possible world to the actual world in which e does not occur is one in which c occurs but simply fails to cause e. Thus, in the case in which an actual c causes an actual e, Lewis thinks that 'if e had not occurred, c would not have occurred' is false; what is true is 'if e had not occurred, then c would have occurred anyway but failed to bring about e'. So '$\sim O(e) \Box\!\!\rightarrow \sim O(c)$' is debarred from being able to account for sufficiency$_c$.

There is, then, nothing in Lewis' account that does or can be used to get at the idea of causal sufficiency$_c$. A lesson in this is that if an event c is necessary$_c$ and sufficient$_c$ for e, then e must be necessary$_c$ and sufficient$_c$ for c, and some alternative method will have to be found for insuring the asymmetry of the causal relation. In counterfactual terms, if c is causally necessary$_c$ and sufficient$_c$ for e, then 'if c had not occurred, e would not have occurred' and 'if e had not occurred, c would not have occurred' are both truths, and some other means will have to be found to show why, when c causes e, e is not also a cause of c.[5,6]

The City University, London © David-Hillel Ruben 1981

---

[4] See for example both his *Counterfactuals* pp. 74–7, and 'Causation', pp. 190–1.

[5] If we accept both these counterfactuals as true, we shall have to reject Lewis' claim that the most similar non-e world in which c occurs but fails to cause e is more similar to the actual world than is the most similar non-e world in which c does not even occur. Lewis argues that both worlds contain one law violation, but that the first differs from the actual world in respect of one particular fact, the second in respect of two particular facts. Swain has an argument against this similarity judgment in the article I have already cited, pp. 8–9.

[6] Thanks to John Mackie and Mark Sainsbury for their comments on and criticism of this paper.

LOCKE ON CAUSATION, COMPATIBILISM AND NEWCOMB'S PROBLEM

By ANDRÉ GALLOIS

DON LOCKE'S reply to my criticisms of his treatment of issues related to Newcomb's problem in 'How to Make a Newcomb's Choice' seems to me to miss the point of those criticisms. No doubt this is because I stated them with insufficient clarity. I will try to do better here.

I

In his original paper Locke dismisses the relevance of reverse causation to Newcomb's problem on the grounds that even if reverse causation is possible, the past is unalterable.

> But even if the actual existence of a Newcomb Predictor were taken as strong evidence for precognition or some other form of reverse causation, it would, presumably, be a matter of the past prediction's being dependent on or determined by the present choice, not a matter of the present choice's actually altering a past prediction, making it other than it was, a suggestion which seems to me to be incoherent (Don Locke 'How to Make a Newcomb Choice' ANALYSIS 38.1, January 1978, p. 18).

In response to this passage I suggested that, if reverse causation is operative in a Newcomb situation, the only clear sense in which the Predictor's past prediction would be unalterable is that the Chooser choosing in one way rather than another cannot bring it about both that the Predictor made a certain prediction and that that prediction was never made. That is, nothing can happen now to bring it about both that some past event occurred and that it did not. However, in this sense present and future events are as unalterable as past events, so the unalterability of the Predictor's prediction would have nothing to do with that prediction lying in the past.

In his reply Locke appears to agree with this point and explains what he has in mind by the unalterability of the past as follows:

> But we can distinguish between altering the present, past or future in the sense of making it other than it actually is, was or will be, which is of course impossible; and altering it in the sense of making it other than it would otherwise be, i.e. without the intervention of the altering event (Don Locke 'Causation, Compatibilism and Newcomb's Problem' ANALYSIS 39.4, October 1979, p. 210).

I am not sure what Locke means by 'altering it in the sense of making it other than it would otherwise be', but if he means that the occurrence of some event E constitutes an alteration of the past just in case there is some event E′ where E′ occurred earlier than E and E′ would not have occurred had E not occurred, then someone who believes that there are cases of reverse causation will believe that it is possible to alter the past in this second sense (had the plane crash not taken place Jones would not have antecedently dreamed about it).

In his original paper it seemed to me that Locke was not simply content to claim that the Predictor's past prediction is unalterable by the Chooser's present choice whether or not reverse causation obtains, but that he was offering the following defence for that claim. Whichever choice the Chooser makes it cannot make a difference to the amount of money in Box I, because either the Predictor has placed a million pounds in Box I or has not placed a million pounds in Box I. Construed as support for, rather than just a restatement of, the claim that the past is unalterable whether or not reverse causation is possible, this struck me as what I called the standard argument for fatalism applied to the past rather than the future. In making this allegation I was suggesting that it is a bad move to defend the unalterability of the past by invokng an argument, the standard argument for fatalism, whose validity is so hotly disputed. At least it is a bad move to do so unless one is prepared to offer a supplementary argument for the validity of the fatalist argument. Consequently it is irrelevant for Locke to remark: 'There is therefore, a certain irony in Gallois' charge that my argument "is simply a standard argument for fatalism applied to the past instead of the future", since the past is precisely where fatalism can be expected to apply.' (ANALYSIS 39.4, p. 210). The point is that, whether or not fatalism applies to the past, Locke has given no argument for saying that it does, other than one whose validity is highly controversial. Hence, so far as I can see, Locke has offered no telling consideration that would preclude us from holding that the past is unalterable just because we are not in a position to causally influence past events, that is, by our present acts cause those events to have happened.

II

In 'How to Make a Newcomb Choice' Locke contended if the Predictor in the Newcomb situation is infallible, then once he has made his prediction it is impossible for the Chooser to make a choice other than that predicted. It was unclear to me what notion of possibility Locke was appealing to here. However, I argued that it would not be causally impossible for the Chooser to have made a choice other than the one actually made even if it were causally impossible for the Predictor to mispredict. In replying to this argument, Locke remarks:

> But this again ignores the unalterability of the past: it may have been causally possible *then* for the Predictor to have acted differently, but it is

> not causally possible *now*, so neither is it causally possible for me to act differently now (ANALYSIS 39.4, p. 211).

It will help to put matters in a clearer perspective if I distinguish two senses in which something may be said to be causally impossible. Let us say that it is causally impossible$_1$ that p if and only if p is logically incompatible with the conjunction of the members of some set of true propositions S, where the truth value of any of the members of S depends only on the obtaining of some causal law, law of nature, natural necessity, empirical regularity or the like. Let us, on the other hand, say that it is causally impossible$_2$ that p if and only if p is logically incompatible with the conjunction of some true proposition whose negation is causally impossible$_1$ together with some true proposition which expresses a hard fact about the past. Finally, following current practice, I will say that a proposition expresses a hard fact about the past if and only if its truth value depends on the obtaining of some past state of affairs and it is logically independent of any proposition whose truth value depends on the obtaining of a present or future state of affairs.

It seems to me that when philosophers claim that something is not causally possible they, generally speaking, mean causally impossible$_1$. Be that as it may, the important question is what Locke means by 'not causally possible'. Locke thinks that if it is not causally possible, in his sense, for the Chooser to have made a different choice, then 'the Chooser cannot choose in any other way'. Certainly if it is causally impossible$_1$ for the Chooser to make a given choice, then that choice is not open to him. However, there are two reasons why Locke, very likely, does not mean by 'not causally possible' causally impossible$_1$. First, if something is ever causally impossible$_1$, it is always causally impossible$_1$ (unless we suppose that it is possible for the laws of nature to change, which is of course a question that is not at issue here). However, in the last passage quoted above, Locke clearly implies that whether or not something is causally possible can vary with time. Second, it is manifestly false that, in a Newcomb situation, it is causally impossible$_1$ for the Chooser to have made a choice other than the one actually made even if it is causally impossible$_1$ for the Predictor to make a mistake. A claim that the Predictor has already made a certain prediction expresses, if true, a hard fact about the past, not the negation of a causal impossibility$_1$. So we cannot move from: it is causally impossible$_1$ that (the Predictor predicts that the Chooser makes choice 1 and the Chooser does not make Choice 1) together with: the Predictor predicts the Chooser makes Choice 1, to: it is causally impossible$_1$ that the Chooser does not make Choice 1. To make this inference would amount to committing a version of what I called in my reply the familiar modal fallacy, where the relevant modality is causal.

Perhaps Locke means by 'not causally possible' causally impossible$_2$.

There is, however, a problem with this interpretation. Though I did not put it in terms of a distinction between two senses of causal impossibility in 'How Not to Make a Newcomb Choice', I suggested in effect that no compatibilist would allow that it follows from a certain choice being causally impossible$_2$ that that choice is not one that is open to the Chooser or not one that the Chooser is free to make. I take compatibilism to be the thesis that there could be actions that agents are free to perform and choices they are free to make which actions and choices are not in fact performed or made, even if determinism is true. By determinism as it applies to acts and choices, I mean the thesis that the occurrence or non-occurrence of any act or choice is at a certain time a logical consequence of the obtaining of natural laws, empirical regularities, etc., together with hard facts about the past relative to that time. So understood it is, of course, an elementary consequence of determinism that non-actual choices and actions are causally impossible$_2$. Indeed, one way to state a standard compatibilist view is as the thesis that an agent may be free to make a choice or perform an action which is causally impossible$_2$. Nevertheless, I have the impression that Locke thinks it ought to worry a compatibilist who supposes that the Chooser is free to make a choice other than the one predicted if it should turn out that it is not causally possible, in Locke's sense, for the Chooser to make such a choice.

The situation appears to be this. If Locke means by 'not causally possible' causally impossible$_1$, then, while a choice that is causally impossible$_1$ is not one that the Chooser is free to make, it does not follow from the Predictor's infallibility that it is causally impossible$_1$ for the Chooser to make a choice other than the one predicted. On the other hand, if Locke means by 'not causally possible', causally impossible$_2$, he owes us an explanation as to why it should be thought to follow from a choice being causally impossible$_2$ that it is one the Chooser is not free to make.[1] Finally, if Locke means by 'not causally possible' something other than causally impossible$_1$ or causally impossible$_2$, I am at a loss to see what it could be.

University of Queensland

© André Gallois 1981

---

[1] It is true that Locke refers to an argument against compatibilism advanced in his article, 'Three Concepts of Free Action,' *Proceedings of the Aristotelian Society*, Supplementary Volume XLIX, 1975. I can find no argument against compatibilism in that article except for one on page 110 which runs:

> But on the other hand, if an action is free in the sense that the agent can act differently from the way in which he actually does act, then it follows that nothing causes him to act as he does. It is, I believe, a necessary—though not a sufficient—condition of an agent's being able to perform a certain action that it is empirically, both physically and psychologically, possible for him to act in that way: if it is empirically impossible for A to do x, then A is unable to do x. Consequently, if the circumstances are such as to make it empirically impossible for A to do x, then in those particular circumstances A

> is unable to do x, whatever might be possible in different circumstances. And finally, on any understanding of causation, if circumstances C are such as to cause E to occur, then the non-occurrence of E given C is an empirical impossibility. This will hold true even if causation is analysed by reference to empirical regularities, since a parallel analysis will also be given for empirical possibility, and it will remain the case that if C causes E then the conjunction of C and not E is an impossibility of the analysed sort. Causal possibility is, after all, a prime form of empirical possibility. Thus a free action, in the sense of one where the agent is able to act differently, will have to be one which lacks a determining cause.

Clearly, Locke's talk of it being empirically impossible for A to do x raises all the difficulties raised by his saying that it is not causally possible for the Chooser to make a certain choice. If he means by 'empirically impossible' causally impossible$_1$, then it does not follow from there being circumstances C that cause an event E to occur that it is empirically impossible for E not to occur. However, if he means by 'empirically impossible' causally impossible$_2$, then it will scarcely be news to a compatibilist that if determinism is true, then if A does not do x, it is empirically impossible for A to do x.

THE INEVITABILITY OF THE STATE

By Diana T. Meyers

ONCE dissected, Robert Nozick's central argument in *Anarchy, State, and Utopia* yields some surprising results. His argument proceeds from the two necessary conditions he gives for the state: (1) it prohibits private enforcement of rights, and (2) it enforces the rights of everyone within its territory (pp. 23–5). In order to show that the state could emerge from a Lockean state of nature without violating anyone's natural rights, Nozick undertakes to explain how an institution meeting these conditions could evolve in a morally unobjectionable way from the system of private protective associations he posits as prevailing in the state of nature.

Nozick's argument can be schematized as follows:

Dominant Protective Association:
  (1) allows private enforcement of rights
  (2) enforces the rights of paying clients

Ultra-Minimal State:
  (3A) prohibits unreliable enforcement procedures and prohibits resistance to its own procedures
  (3B) prohibits private enforcement of rights
  (2) enforces the rights of paying clients

Minimal State:
  (3B) prohibits private enforcement of rights
  (4A) enforces the rights of non-clients whom it prohibits from enforcing their own rights
  (4B) enforces everyone's rights

A justification of the ultra-minimal state's consolidation of enforcement power requires a demonstration that the move from 1 to 3A would not violate anyone's rights and that 3A is equivalent to 3B. Then Nozick must show that the substitution of 4A for 2 would infringe no one's rights and that 4A and 4B are equivalent. This second line of argument legitimates the transformation of the ultra-minimal state into the minimal state. In this paper, I will focus on the first step of each argument, from 1 to 3A and from 2 to 4A.

Properly understood, these two arguments are independent. If the second is built upon the first, it begs the question. For the assumptions that make Nozick's first argument forceful make his second argument trivial. Yet, if, instead, assumptions appropriate to the second argument are introduced into the first, the first argument is reduced to an unpersuasive tautology. Thus, it is only if a different set of assumptions is adopted for the purposes of each of the two arguments that they can carry their separate weight.

In particular, Nozick contends that his argument for the ultra-minimal state does not depend on the assumption that persons have rights specifying what procedures must be followed in determining their guilt or innocence when accused of crimes. If persons have procedural rights, the right to punish is constrained by them, i.e., it is a right to punish persons reliably found to be guilty of crimes. This granted, independents who decline to purchase the dominant protective association's services and favour unreliable enforcement procedures are not exercising any natural right. They are arbitrarily violating others' rights while rationalizing their conduct by invoking the right to punish. Since they are not exercising a natural right, there is nothing to prevent the dominant protective association from barring their pretended enforcement activities. Once procedural rights have been admitted, moral obstacles to the move from 1 to 3A dissolve.

Nozick, however, recognizes that procedural rights take him too directly to his conclusion. Thus, he urges that the ultra-minimal state can be legitimately attained without the 'facilitating assumption of procedural rights' (pp. 106–7). Let us suppose that Nozick has improved on Locke's glib 'I easily grant that civil government is the proper remedy for the inconveniences of the state of nature' by showing that the appeal to procedural rights is unnecessary for purposes of justifying the ultra-minimal state. How does this affect his argument for the moral legitimacy of the minimal state?

The minimal state not only prohibits private rights enforcement but also protects those whose enforcement activities it has prohibited. Nozick claims that the ultra-minimal state is morally bound to transform itself into a minimal state because it is obligated to compensate repressed independent enforcers and protection is the obvious remedy (p. 110).

This result follows immediately from the supposition that the right to punish is not constrained by procedural rights. If independents using unreliable procedures are acting within their rights, the ultra-minimal state must compensate them for its infringement of their rights, and rights enforcement is the ultra-minimal state's stock-in-trade.

At this point, it might be objected that since we have procedural rights (Nozick apparently thinks we do), there is no need to compensate independents when their procedures are proscribed, for no right of theirs is being infringed. Unless the right to punish is unrestricted, compensation is superfluous. From the standpoint of the move from 2 to 4A, the denial of procedural rights proves to be the strong assumption. Here, Nozick allows that compensation must be paid for the wrong inflicted on independents and therefore that the minimal state's redistributive enterprise is morally required, when it has not been established that the ultra-minimal state has committed any injustice.

The assumption that persons have no procedural rights gives Nozick's argument for the ultra-minimal state whatever power it has but trivializes his argument for the minimal state. On the other hand, the contrary assumption, namely, that persons have procedural rights, while capable of lending force to an argument for the minimal state, renders the argument for the ultra-minimal state tautological. What lesson can be drawn from this?

It would be a mistake for anarchists to applaud the present reconstruction of Nozick's argument. The explication I have given might tempt us to conclude that the idea of the state is self-contradictory on the grounds that its two components can only be deduced from incompatible premisses. However, I have only maintained that the most convincing arguments for Nozick's necessary conditions stem from incompatible assumptions. In *Anarchy, State, and Utopia* Nozick contends that both necessary conditions can be obtained by arguing from the assumption that persons have no natural procedural rights, and it is open to him to show that both are deducible on the assumption that persons have natural procedural rights, too.

Instead of capitulating to the anarchists, Nozick should concede that Locke was right to treat the moral legitimacy of the state, though not the moral legitimacy of each particular state, as virtually self-evident. For the right to punish, whether it is conceived as a carte blanche or as an option narrowly circumscribed by others' interests in their security, generates compelling reasons for instituting the state. Formulating this in a more Lockean manner, we may depict victims of crime with their right to punish and defendants with their right to be judged in accordance with reliable procedures as natural enemies. They are inevitably at odds and in need of an authority to monitor their exercise of these rights. Once the need for such an authority is acknowledged, nothing stands in the way of that authority's assuming the powers of an agency to exercise these rights on behalf of individuals. It is, after all, the authority that is in doubt, not the power.

Perhaps Nozick regards the state as more problematic than it is because he thinks that natural rights never conflict, each person's rights being limited by the rights of others (cf. pp. 173, 180–1, and 238). If this view is supplemented with a belief in human omniscience regarding the range of morally available alternatives in varying circumstances, there is no reason to conclude that the state is necessary. Unfortunately, it is not clear that natural rights can be analysed so as never to conflict, but it is clear, in any event, that people lack the requisite moral insight to take advantage of such theoretically harmonious natural rights. Consequently, conflict between right-holders is in principle unavoidable and so is the state.

Cornell University

© Diana T. Meyers 1981

NOZICK ON HART ON THE RIGHT TO ENFORCE

By HILLEL STEINER

IN *Anarchy, State and Utopia,* (Oxford, 1974), Robert Nozick challenges a claim advanced by H. L. A. Hart, that having a right entails having a right to enforce fulfilment of the obligation correlative to that (former) right.

> If I cautiously insist that you first promise to me that you won't force me to do A before I will make my promise to you to do A, and I do receive this promise from you first, it would be implausible to say that in promising I give you the right to force me to do A. (p. 92)

Nozick does not here explicitly say that the claim of the recipient of the promise to do A actually amounts to a right. But it seems fair to assume that he does so consider it, since otherwise his counter-example could not serve its intended purpose of disproving Hart's view that having a right to A entails having a right (strictly, a power) to enforce the doing of A.

I think Nozick's argument is mistaken. His example looks to be analogous to the case of a woman who, as a condition of marriage, extracts from the man in question an undertaking that he will remain a bachelor. Nozick is claiming that Hart's right that Nozick do the promised A does not entail Hart's right that Hart force Nozick to do A, because Hart has promised not to force Nozick to do A. But why, we may ask, does Hart's promise not to force imply Hart's lack of a right to force? Presumably it has this effect because, like Nozick's promise to do A, it places the promisor under an obligation to the promisee. In this case, Hart is presumed to have an obligation to Nozick not to force. Hence, and again as with Nozick's promise to Hart, the promisee acquires a right: Nozick must be said to have a right that Hart not force.

Now, does this right that Hart not force entail, in Nozick, a further right (power) to force Hart not to force? On his general thesis—that a right to something does *not* entail a right to enforce the correlative obligation—Nozick is committed to denying that he himself has this right. Yet if Nozick lacks this further right, Hart must lack its correlative liability which is one to be subject to an obligation not to force. But the implied effect of Hart's promise is that it places him, Hart, under just such an obligation. So Nozick's argument is self-contradictory since it implies at once that Hart both is, and cannot be placed, under an obligation not to force.

University of Manchester

© HILLEL STEINER 1981

ISOLATING INTRINSIC VALUE

By Eva Bodanszky *and* Earl Conee

RODERICK CHISHOLM has recently proposed analyses of the basic concepts of intrinsic value. ('Intrinsic Value', *Values and Morals*, ed. A. I. Goldman and J. Kim, Reidel 1978, pp. 121–30). He derives his approach from the view of Aristotle and Moore that the intrinsic value of something is the value that it "would have in isolation". Chisholm seeks to defend this view by offering precise analyses which incorporate the intuition that an intrinsically good state of affairs is one such that 'its goodness doesn't require that there obtain some *other* good state of affairs which neither includes nor is included within it' (ibid. pp. 121–2). We shall try to show here that Chisholm's proposed analyses do not succeed. We maintain that they fail in ways that indicate that nothing like his isolation approach yields adequate explanations of these concepts.

Chisholm's account appeals to a single normative primitive: the concept of being "generically better". No explanation of this notion is given. The way it is used in the exposition of the account requires that it establish a value ranking in which a state of affairs gets credit for supplying either intrinsic or extrinsic benefits, and suffers discredit for having either intrinsic or extrinsic liabilities. For example, something intrinsically neutral is generically better than neutral if it has good effects, and something intrinsically good is generically neutral if the disvalue of its effects is equal to its intrinsic value (see ibid. pp. 121, 122, 124). Appealing to this notion of being generically better, Chisholm constructs his analysis of being intrinsically better, and uses that to explain intrinsic goodness and evil. This is the analysis:

> D1 *p is intrinsically better than q* $=_{\mathrm{df}}$ p and q are bearers of intrinsic value; and any world, W1, such that p reflects all the good and evil there is in W1 is [generically] better than any world, W2, such that q reflects all the good and evil there is in W2 (ibid p. 126).

The technical terms of D1 are defined as follows:

> D2 *p reflects all the good and evil there is in q* $=_{\mathrm{df}}$ q entails p; and every part of q that entails p has the same [generic] value as p (ibid p. 123).

The relevant concepts of entailment and parthood are these: 'p entails q $=_{\mathrm{df}}$ p is necessarily such that (i) if it obtains then q obtains and (ii) whoever accepts it accepts q, . . . a state of affairs p is a "part" of a state of affairs q, if and only if q entails p' (ibid.).

D3 *p is an intrinsic value state* $=_{df}$ there is a possible world, W, such that: p reflects all the good and evil there is in W; and if p is not [generically] neutral, then everything that reflects all the good and evil there is in W entails p or is entailed by p (ibid. p. 124).

D4 *p is a bearer of intrinsic value* $=_{df}$ p is an intrinsic value state and not-p is an intrinsic value state (ibid. p. 125).

Unfortunately, even the elaborate conceptual machinery of D1 does not suffice to "isolate" the intrinsic goods. They retain logical and causal accompaniments that ruin the account. In the first place, most of the examples of intrinsic goods that Chisholm uses in developing the analyses fail to be bearers of intrinsic value, although by D1 only bearers of intrinsic value are intrinsically better than anything. Consider:

(a) Brown experiencing two units of pleasure.

For (a) to be a bearer of intrinsic value, it must be an intrinsic value state. It is not. The worlds where (a) reflects all the good and evil must be worlds in which the only thing that independently contributes value is Brown's pleasure, if D1 is to work properly. But in any such world there are many logically independent states of affairs which also reflect all the good and evil, for instance either

(b) Someone experiencing two units of pleasure while wearing a hat,

or

(b′) Someone experiencing two units of pleasure while not wearing a hat.

So each such world contains a part (either (b) or (b′)) which reflects all the good and evil there, but neither entails nor is entailed by (a). Since (a) is non-neutral in these worlds, by D3 it is not an intrinsic value state. Thus D4 tells us that it is not a bearer of intrinsic value, and so not counted by D1 as intrinsically better than anything.

The causal problem for D1 is made by the fact that intrinsically good states of affairs that can have bad effects reflect all the good and evil in some generically neutral worlds. Because of the way D1 evaluates states by use of such worlds, these fail to be rated intrinsically better than neutral (Chisholm was aware that this problem might arise, and sought to avoid it—see ibid. p. 124; we argue that the problem remains). In the case of (a), this can be shown as follows. There is a possible world, Wa, where (a) causes

(c) Brown experiencing two units of *dis*pleasure, and those two experiences are the only two things that independently contribute value to the world. Nonetheless, (a) reflects all the good and evil in Wa. Since (a) causes the intrinsically evil (c), it has neutral generic value in Wa. The other parts of Wa that entail (a) (including Wa itself) are conjunctions of (a) with the other truths of Wa. When such a part also includes (c) as a conjunct, it is clearly intrinsically and generically neutral. But even when such a part does not include (c) it is generically neutral, for it is (a)'s *generic* value in Wa that determines its contribution to the generic value of these conjunctions. To see this, consider what would be relevant to the evaluation of such a conjunction if it were the only known consequence of a contemplated act. We would want to know its generic value, since we would be concerned with both its intrinsic and extrinsic contributions. And it would be a clear mistake to ignore (a)'s extrinsic disvalue in making this assessment. That would rate too highly the act of bringing about the conjunction. Thus the generic value of a conjunction depends upon that of its conjuncts. So the parts of Wa that entail (a) are all generically neutral. Since Wa is a world where (a) reflects all the good and evil, according to D1 (a) is not intrinsically better than intrinsically neutral states of affairs.

We believe that D1 is beyond repair. And we think that its problems typify the difficulties that face attempts to characterise intrinsic betterness by use of generic betterness and some sort of isolation. In particular, reflection upon these problems makes it seem quite doubtful that there is any sort of situation where each state of affairs has exactly its intrinsic value as its generic value. And without that there appears to be no hope for the approach.

© EVA BODANSZKY *and* EARL CONEE 1981

University of Massachusetts, Amherst, and
Indiana University

IMPERATIVE INFERENCE—AN ADDENDUM

By VERA PEETZ

I AM sorry that Professor Hare (ANALYSIS, 39.4, October 1979, p. 161, footnote 1) thinks that I misrepresented him when I said ('Imperative inference', ANALYSIS, 39.2, March 1979) that he had suggested that there are imperative inferences whose premisses are all indicatives. The precise reference in *The Language of Morals*, which as Hare rightly says I should have given, is page 33, line 25, where Hare says '. . . there is one kind of imperative conclusion which can be entailed by a set of purely indicative premisses.' Now, of course, Hare is here talking about hypothetical imperatives which he regards as somewhat different from the ordinary imperatives which he had discussed in the previous chapter, but, from my point of view, what I say about imperative inference will also include these hypothetical imperatives. I did mention that Hare had given rules for imperative inference, but I did not say what these rules were and I am sorry if I thereby gave an unbalanced view of what Hare had said.

What I should have said, but did not, is that Hare's rules for imperative inference, namely

(1) No indicative conclusion can be validly drawn from a set of premisses which cannot be validly drawn from the indicatives among them alone.
(2) No imperative conclusion can be validly drawn from a set of premisses which does not contain at least one imperative,

will apply in a modified form to inferences containing descriptions of imperatives. The modified form will be something like

(1′) No indicative conclusion which does not contain a description of an imperative can be validly drawn from a set of premisses which cannot be validly drawn from the indicatives among them, which do not contain descriptions of imperatives, alone.
(2′) No conclusion describing an imperative can be validly drawn from a set of premisses which does not contain at least one description of an imperative.

The modified form of these rules will apply also to *ought* if it is regarded as descriptive of a prescription and that they do apply is at least part of what is meant by saying that an *ought* cannot be derived from an *is*. I would like to suggest that such an *ought* is sufficient for Hare's needs. As Hare says, the use of a specific *ought* implies or presupposes a universal *ought*, so that when one says 'You ought to do X', this presupposes an argument such as

Everyone ought always to do X in circumstances C
You are now placed in circumstances C
Therefore, you ought to do X now

(where *ought* is being used in a descriptive sense). Now according to Grice's conversational maxim (H. P. Grice, 'Logic and Conversation' in *Syntax and Semantics* Vol. 9, ed. P. Cole, Academic Press, N.Y., 1978) that one ought not to say more or less than one needs to, it would seem that, if the use of *ought* presupposes an argument such as the one above, then 'You ought to do X' conversationally implies that the speaker subscribes to the prescription, although not actually uttering a prescription himself. If he wants to show that he does not subscribe to the prescription, then he has to deny it specifically: 'You ought to do X, but do not do X'. So, while Hare's rules (in their modified form) are necessary for inferences involving descriptions of imperatives (including *oughts*), Hare does not need inferences involving explicit imperatives for his moral theory.

University of Nottingham

© Vera Peetz 1981

IMAGINATION, DESIRE AND PRESCRIPTION

By H. M. Robinson

In his paper 'Relevance' (A. I. Goldman and J. Kim edd., *Values and Morals*, Reidel 1978, pp. 73–80), R. M. Hare returns to the knotty problem of trying to show how the principle of universalizability rules out racism. He argues that if the white racist imagines himself in the position of his black victim, assuming the desires of the latter, and thereby discovering that the black man's desire not to be discriminated against is stronger than his desire to discriminate, then he can no longer sincerely prescribe racism. He argues this on the grounds that once he has acquired, in imagination, the black man's desire

> which *ex hypothesi* is greater than his original desire to do the act, he cannot, if he is to make the same judgement about the actual situation as about the hypothetical similar one (another logical requirement), give greater weight to his own original desire than to his victim's; for to do so would be to give greater weight to the lesser desire among desires

> *which he himself has* (one of them originally, and one of them latterly in his imagination). That is why he cannot really and sincerely assent to the principle on which he claims to be relying, that whites ought to do this to blacks in such a situation. (pp. 88–9, italics in original.)

And also

> . . . if the relative strengths of the desires is as we have supposed, they cannot then go on appealing to these racist principles, because they cannot desire, and therefore cannot sincerely prescribe, that the greater desires of theirs should be subordinate to lesser (p. 89).

Thus Hare is grounding his position on the very strong claim that the universalizing racist is actually acting against his own desires. He is thus not prescribing what he himself really wants. His prescription does not merely conflict with the desires of others, or with some formal principle of ethics (e.g. universalizability) but with his own desires. Though Hare does not say so here, that would actually make the prescription imprudent according to the standards of prudence he adopts elsewhere (e.g. 'Ethical Theory and Utilitarianism' in *Contemporary British Philosophy*, ed. H. D. Lewis, London, George Allen and Unwin 1976, p. 116 and p. 119). He does say here that it makes it insincere.

The view expressed in these quotations is not novel. In *Freedom and Reason* Hare asserted that a man cannot sincerely act against hypothetical desires if they are appropriately strong (pp. 108–9). The view is vital to Hare's moral psychology, for it explains why the moral viewpoint is the rational one, as follows. Anyone who wants to be fully informed of the facts will want to know how others are affected by actions: this, Hare says, involves acquiring their desires and once one has done this it is rational—i.e. a function of one's own desires— to act in the way that maximizes satisfaction overall. Hence a serious problem inherent in previous forms of non-cognitivism is overcome; for whereas emotivists had conceded that once agreement on factual matters was reached there remained the possibility of sincere divergence of moral attitude, on Hare's position this is not possible, for it is part of being fully informed that one acquire all these attitudes and treat them as one's own. (Vide 'What Makes Choice Rational?' *Review of Metaphysics*, vol. 32, 1979, pp. 623–37. In this paper Hare argues that being fully informed eliminates all subjectivity from prescriptivism. See especially pp. 636–7.)

It is essential to the generation of this argument that it be the case that someone who takes on a desire imaginatively can be said to have that desire himself. But this is most implausible. If the racist (call him White) really enters into his victim's position the following propositions will be true of him.

(1) White realizes that if he were black he would desire not to be treated as inferior.

(2) White now has the desire that if he were black he be not treated as inferior.

But there seems no way of deriving from (1) and (2)

(3) White now has the desire that blacks be not treated as inferior.

It seems so clear that imaginative substitution can only generate the likes of (1) and (2) and that they do not entail (3) that it is difficult to see how Hare comes to assert (3). *Perhaps* he thinks that what is derived from imaginative substitution can be expressed as

(4) White now desires that he, *considered qua black*, be not treated as inferior.

It seems correct that 'x, *qua* F, is G' entails 'x is G'. However, in this case (4) means nothing unless it is a clumsy way of stating (2), for something cannot be considered transparently *qua* something it is not. 'x, *qua* F, *would be* G' does not entail 'x is G'.

It is difficult to avoid the conclusion that Hare treats having a desire *in imagination* as a manner of actually having the desire. But having a desire in imagination is imagining (what it would be like) having the desire, which is not a manner of having it. Rather it can only give rise to something like (1) or (2). It is, therefore, difficult to see why White should not sincerely affirm (1) and (2) whilst denying (3). He cannot therefore be obliged to act on (3).

Of course, if the principle of universalizability itself directly requires that a moral agent give equal and impartial weight to everyone's desires, then racism will be ruled out. But it seems to be Hare's strategy to derive this impartiality indirectly, *via* the imaginative substitution which, he holds, universalizability directly commands. If this were not his strategy it would be difficult to see why Hare should put such weight on this imaginative exercise, for it would otherwise have only a marginal value as a technique for discovering just what the strength of others' desires probably are.

There are suggestions elsewhere in Hare that he is not too clear about the relation between the imaginative substitution and universalizability. In his review of Rawls he says

> My own theory secures impartiality by a combination of the requirement that moral judgements be universalizable and the requirement to prescribe for hypothetical reversed role situations as if they were actual (I am not sure whether the second is an independent condition or not) (*Philosophical Quarterly*, vol. 23, 1973, p. 154).

My object so far has been to show that the latter requirement cannot be based on imaginative substitution and therefore it is to be hoped that it can be derived more directly from universalizability. As a natural correlate I shall argue that if universalizability can perform this role no central purpose is served by imaginative substitution.

II

Once we have shown that imaginative substitution does not imply *having* the desires of all involved, then it follows that sincerity, considered as fidelity to one's own desires, cannot be used to generate impartiality.

Hare's notion that one actually has all the desires into which one enters imaginatively explains what would otherwise appear to be a bizarre use of 'sincerity'. As the moral agent imaginatively substitutes himself in each position, each gives him a comparison with his own point of view and he can be, or fail to be, sincere about what he would want (and how strongly) in *that* position. But this is an obviously insufficient basis for action: one cannot say that if someone cannot sincerely prescribe some principle from every viewpoint involved, then he cannot sincerely prescribe it *simpliciter*, for one could never prescribe punishment because, from the victim's point of view, one could not sincerely desire it. On the other hand the idea that he can be sincere about what he would want if he were in all positions at once is incoherent, for the idea that he could be in all positions at once is itself incoherent (or will usually be so). However, if Hare thinks that imagining oneself in the position gives one simultaneously the desires associated with each position, that would explain how sincerity could take the full scope of all positions, for, though one is not actually in all the positions, nevertheless one's psychology is a sum of all of them and therefore the test of sincerity becomes applicable. But if one accepts that one cannot be in all positions and that the psychological state which follows from imaginative substitution is not equivalent to having all the desires, then sincerity is irrelevant. What imaginative substitution gives one is, at most, knowledge of the desires of all parties and a basis for comparing and computing their desires and how to maximize satisfaction overall.

The process of deciding what is the most utile action has two phases. First, one must decide what desires each party has and how strong they are. Second one must compare and compute those to see what action will maximize satisfaction. The arguments above show that imaginative substitution does not necessarily lead one to do the latter on pain of insincerity. Nor does such imagining seem to be necessary for such computing, for once one knows the strengths of all the relevant desires the computation can presumably be done in a purely mechanical way. Does imaginative substitution perhaps have an essential role in the first task, that of assessing the nature and strengths of the relevant desires? Hare clearly thinks that it does for he says

> the agent does not have . . . full knowledge and awareness of one of the available facts, namely, the sufferings of his victim, unless he is aware of what it is like to be that person; and this he cannot be aware of unless he puts himself in thought in that person's position ('Relevance' p. 81).

However, it does not seem that imagination can have such a vital role, because unless one can in principle work out independently what someone is desiring or feeling it is impossible to imagine what it feels like. For example, I cannot work out what it would be like to have a Martian's sixth sense by imagining myself in his position, rather I must know what it is like to have that sense before I can imagine his position. The imaginative substitution can be used only to *bring home* something that I can intellectually know otherwise. If someone has a type of pain that is beyond my experince (for example, some psychological affliction of an unusual sort) I have a better chance of estimating from his behaviour how unpleasant it must be than I have of imagining just what it must be like from trying to assume his position. Even in a more commonplace case, when I think I can imagine what I would feel in the other's place, the only reason I could have for thinking that this approximates to how he does actually feel (and if it does not how can my feeling serve utility?) is if I have taken care to imagine those feelings and desires which fit with his actual behaviour. Thus public evidence, such as behaviour, must be more fundamental than imagination.

I therefore find it difficult to see how imaginative substitution plays any vital role. One can just as well or better infer from behaviour what desires someone has as one can imagine those desires; and once one has in some sense grasped the desires the obligation to compute them (if there is such an obligation) must follow from some abstract principle (e.g. strong universalizability) for it cannot be derived, via imagining, from the need to be faithful to one's own desires.

University of Liverpool © H. M. Robinson 1981

MILL, PATERNALISM, AND SLAVERY

By JOHN D. HODSON

AT one point in *On Liberty*, Mill directs his attention to the question of how the law should deal with slavery:

> In this and most other civilized countries . . . an engagement by which a person should sell himself, or allow himself to be sold, as a slave would be null and void, neither enforced by law nor by opinion. The ground for thus limiting his power of voluntarily disposing of his own lot in life is apparent, and is very clearly seen in this extreme case. (ch. V)

Mill proceeds to defend the nonenforcement of slavery contracts on the ground that this interference with a person's liberty is for the sake of that same person's liberty; respect for liberty does not require that we allow a person to abdicate liberty. This position of Mill's has been taken to be an exception to his well-known opposition to strong paternalism.[1] Those unwilling to embrace strong paternalism have been forced to seek other justifications for the law's refusal to enforce slavery contracts.[2] Without suggesting that these approaches are unsuccessful, I want to suggest that they are unnecessary. They are unnecessary because, as I shall argue, *there is no need to justify the refusal to enforce slavery contracts.*

Under what circumstances is it necessary to justify an action or a governmental policy? Moral justification is necessary only when there is some *prima facie* objection to the action or policy. Unless there is some *prima facie* objection, the action or policy is either morally neutral or morally desirable; in those cases no justification is necessary.

I shall argue that there is no *prima facie* objection to the nonenforcement of slavery contracts, and thus that we need neither seek paternalistic justifications nor alternative justifications for such nonenforcement.

Mill and, I suspect, others who have sought to justify the refusal to enforce slavery contracts have found the *prima facie* objection in the assumption that nonenforcement constitutes an interference with the liberty of the would-be slave (henceforth: 'slave'). Accordingly, in the above quotation we find Mill speaking of the 'ground for *thus* limiting [the slave's] power of voluntarily disposing of his own lot in life.' This, however, is a mistake. The refusal to enforce slavery contracts does not constitute an interference with the slave's liberty.

---

[1] See, e.g., Gerald Dworkin, 'Paternalism,' in *Morality and the Law*, ed. Richard A. Wasserstrom (Belmont, Ca.: Wadsworth, 1971), p. 117; Joel Feinberg, 'Legal Paternalism,' *Canadian Journal of Philosophy* 1 (1971), pp. 116–117. Strong paternalism is the view that 'the state is justified in protecting a person, against his will, from the harmful consequences even of his fully voluntary choices and undertakings.' (Feinberg, p. 124.)

[2] See, e.g., Feinberg, pp. 118–20; Tom L. Beauchamp, 'Paternalism and Biobehavioral Control,' *The Monist* 60 (1977), p. 74; Michael Bayles, *Principles of Legislation* (Detroit: Wayne State University Press, 1978), p. 130.

In the first place, the law's refusal to enforce slavery contracts does not prevent anyone from living in *de facto* slavery. One person may submit himself to the will of another in all of the ways associated with slavery; he may act only upon the command or with the permission of another and his life may be devoted to doing the bidding of this other person. None of this is prohibited in a policy of nonenforcement of the slavery contract; what is denied is merely the *de jure* recognition of the slave's status as a slave.

Secondly, then, does the refusal to give legal recognition to the status 'slave' constitute an interference with the slave's liberty? It is not a restriction of his 'power to dispose of *his own* lot in life,' for there is nothing which is his which is thus denied him. Further, there are innumerable statuses which are not recognized in, for instance, the United States, yet no one complains that this is a limitation on liberty. Americans are not able to gain legal recognition as knights, dukes, barons, or princes, but surely the fact that American law does not make these titles officially available is no restraint on liberty. Thus, it is difficult to see how the law's failure to recognize slavery as a legal status interferes with anyone's liberty.

Third, it may be argued that the interference lies in the law's failure to put its coercive authority behind the slavery contract. There are two cases to consider in this regard. One is the case in which *the slave* decides not to fulfil his part of the bargain; the slave wishes no longer to be a slave. However, it can hardly be considered an interference with the *slave's* liberty not to force him to continue to be a slave; enforcement of the contract, not nonenforcement, would constitute the restraint on liberty in this instance.

The other case is that in which it is *the buyer* of the slave who fails to do what he has contracted. Is it an interference with the slave's liberty for the law not to require the buyer to perform his part of the bargain? Note that the law does not require the slave to remain a slave, so the problem here is simply that the law fails to make available its enforcement services. But it is not an interference with individual liberty for the law to fail to make available every service which it could make available. If there are only *negative* rights to liberty, then respect for liberty does not require provision of *any* services; even if there are some positive rights, there is no reason to think that individuals have positive rights to *this* service. Nonenforcement of slavery contracts is thus simply a failure to provide a nonessential service which could be provided; it is not an interference with the slave's liberty.

The traditional *prima facie* objection to nonenforcement of slavery contracts was that this interferes with the slave's liberty. Since this objection now appears unsound, we may consider whether there are any other *prima facie* objections which would make it necessary to justify the position on slavery.

One possibility is that nonenforcement interferes with the liberty of *the buyer* of the slave rather than the slave himself. The buyer is denied the opportunity to enter into certain kinds of relationships, and this might be construed as an interference with liberty. However, this argument fails for the same reasons as the argument that nonenforcement interferes with the slave's liberty. The law's failure to make its enforcement services available in this sphere is simply a failure to provide a service, and the failure to provide a service is not *per se* objectionable.

The fact that the slave in question gives (*ex hypothesi*) his voluntary informed consent to the transaction might be thought to be a significant factor. In other areas we expect someone who undertakes an agreement to perform as promised, especially when his part of the bargain involves no unforeseen costs. A person must take responsibility for his own actions. However, this argument is irrelevant to the question at issue. At best, it shows that it would be *permissible* for the law to enforce the slavery contract; it does nothing to show that there is any *obligation* or *duty* to do so. What is needed is the stronger claim that nonenforcement violates some *prima facie* moral *requirement.*

The final argument I shall consider is this. Other kinds of contracts *are* enforced when the contracting parties knowingly consent to the bargain. The law does provide this service even if it is not morally required to do so. Thus, it may seem that fairness or consistency would dictate that slavery contracts should be enforced in the same manner as other contracts. I submit, however, that the force of this argument depends upon whether the law has promulgated its position on slavery contracts. If it has made this policy public, then no one has the right to expect that slavery contracts will be enforced. The matter is then again simply one of failing to provide a service. It would be unfair or inconsistent not to provide this service only if persons had legitimate expectations that it would be provided; such expectations are eliminated by promulgation of the policy against enforcement.

Unable to discover any sound *prima facie* objections to the law's refusal to enforce slavery contracts, I conclude that there is no need to justify that position. One implication is that the question of slavery contracts provides no reason even to be tempted, as Mill was, by strong paternalism.

The University of Texas at Austin

© JOHN D. HODSON 1981

ON A CLAIM OF HARE'S

By P. T. GEACH

IN a recent article Professor R. M. Hare claims that a view he rejects 'can be shown to be a mistake by giving a completely formal account of the logic of the moral concepts which does not involve any substantial moral commitment'. He at once adds 'obviously this is not the place to attempt such an account' ('Justice and Equality': *Dialectics and Humanism,* vol. VI no. 4, Warsaw, 1979): quite properly, since his paper appears in the proceedings of an Anglo-Polish seminar on justice and equality. However, his claim is that something *can* be done formally, in the domain of logic; and such a claim ought to be substantiated by citation of work that has actually been done, or at least such a citation ought to be producible on demand. If a logician claims that a way of showing schemata of a certain class to be valid or invalid is in fact a decision procedure for the whole class, then he ought either to prove this himself, or to be ready with citation of work on the decidability of theories (see the preface to my *Reason and Argument*, second edition). Similarly for Hare's claim.

I do not think Hare would say that the work he himself has so far published on imperative and deontic logic is itself the 'completely formal account' he is speaking of. His work on imperative logic has been much criticized since he published it in *The Language of Morals* in 1952; and though he has often asserted that the deontic theses that are relevant to morals entail imperatives, he has nowhere shown us how a mixed logic of imperatives and deontic theses would actually work. So I do not think he would or could claim that he has shown in essentials how the 'completely formal account' would work, leaving the details as, so to speak, an exercise for the reader. Other people have done work on imperative and deontic logic, and it is possible that Hare regards some of this work as giving that 'completely formal account of the logic of the moral concepts' which he is alluding to. If so, it is proper to ask whose work is in question and where it it is accessible.

University of Leeds

© P. T. GEACH 1980

ON A MISUNDERSTANDING OF GEACH'S

By R. M. HARE

SNIPING demands accuracy. From the passage which Geach quotes ('On a Claim of Hare's', ANALYSIS this issue; *Dialectics and Humanism*, vol. VI no. 4, Warsaw, 1979, p. 22), and still more from the context which he does not, it will be clear to most readers that he has missed the target. I meant by 'formal', not the opposite of 'informal' (as in the title of Prior's *Formal Logic*), but the opposite of 'substantial'. As regards formal deontic logic and the logic of imperatives, I remain dissatisfied with any contributions I may have made, and also with other work that I know of; the subject has made some progress, but awaits its Aristotle. I think it *can* be developed into a satisfactory system, and no doubt will be; but not by me, and not yet. To say this is not to deny the importance of some recent work in the field, e.g. Professor Rabinowicz's *Universalizability* which, so far as I understand it, could form a part of the sort of system that Geach asks for.

My point was rather that even an informal (no *p*s and *q*s) account of the logical properties of the words used in moral discourse can be formal in the sense of not involving substantial moral commitments (just as an account of the logical properties of 'all' does not involve commitments of substance to the truth of any synthetic statements containing it). I have tried to give such an account in my books and shall develop this further in a new book, *Moral Thinking: Its Levels, Method and Point,* which I hope to bring out before long. Though not departing from formality in the sense explained (no substantial commitments), I attempt to show that useful results can be reached for practical moral reasoning, which shed light on some important current issues. I am not aiming to provide a fully formalized (in the Prior sense) system of deontic logic.

Corpus Christi College, Oxford

© R. M. HARE 1981

Philosophical Books

***Philosophical Books** aims to provide prompt, scholarly reviews of professional books in philosophy. Besides conventional reviews in both contemporary philosophy and the history of philosophy, it now includes regular Discussion features, where the authors of selected titles have the opportunity to reply to reviewers. There are occasional subject surveys and issues devoted to a single theme.*

Books for review should be sent to the Editor, Dr M.A. Stewart, Department of Philosophy, University of Lancaster, Lancaster LA1 4YT.

Subscription prices are as follows:

Institutions:			
£14.50 (UK)	*£17.75 (overseas)*	*$36.00 (USA)*	*$41.50 (Canada)*
Individuals:			
£9.50 (UK)	*£11.50 (overseas)*	*$23.50 (USA)*	*$27.00 (Canada)*
Single issues:			
£4.00 (UK)	*£5.00 (overseas)*	*$10.00 (USA)*	*$11.50 (Canada)*

*There are now **four** issues a year.*
Single copies and back issues are available.

BASIL BLACKWELL PUBLISHER
Journals Department, 108 Cowley Road, Oxford OX4 1JF

RATIO

edited by
MARTIN HOLLIS

Volume XXII *December 1980* *No. 2*

CONTENTS

PETER ALEXANDER	*Locke on Substance-in-General I*
JOSHUA HOFFMAN	*Locke on Whether a Thing can have Two Beginnings of Existence*
CRISPIN WRIGHT	*Realism, Truth-Value Links, Other Minds and the Past*
CHRISTOPHER CHERRY	*Knowing, Imagining and Sympathizing*
WOLFGANG G. STOCK	*The Informatics of Science Establishment, Subject Matter and Methods*
I. A. OMER	*Minimal Law Explanation*
SHALOM LAPPIN	*Why One Ought to Care about Identity across Possible Worlds in Moral Judgments: Reply to Freundlich*

Annual Subscription: £15.25 (inland) £18.75 (overseas) $41.25 (USA) $47.75 (Canada)
Orders to: Journals Department, Basil Blackwell Publisher Ltd.,
108 Cowley Road, Oxford OX4 1JF, England.

The German edition of *Ratio* is published by FELIX MEINER VERLAG,
Richardstrasse 47, 2000 Hamburg 76, West Germany.

Published by BASIL BLACKWELL, Oxford

NOTES

The Analysis Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

Subscriptions. The subscription to Analysis for institutions is £8.00 (inland), £9.80 (overseas), $24.00 (U.S.A.), $28.00 (Canada); for individuals £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A.), $17.50 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

Contributions. **Articles submitted for publication should be addressed to Mr. Christopher Kirwan, Exeter College, Oxford OX1 3DP.**

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

ISSN 0003–2638

PRINTED IN GREAT BRITAIN BY BURGESS & SON (ABINGDON) LTD., ABINGDON, OXFORDSHIRE

Vol. 41 No. 2 (New Series No. 190) March 1981

ANALYSIS

Edited by
Christopher Kirwan

CONTENTS

Begging the question 1981 — Richard Robinson
What is wrong with disjunctive syllogism? — Stephen Read
Kripke's speaker's reference — Rod Bertolet
The discreet charm of counterpart theory — Graeme Hunter *and* William Seager
Actions, reasons and Humean causes — Peter H. Hess
On Mackie's new account of causal priority — Douglas Ehring
Causal priority and the direction of conditionality — J. L. Mackie
Backwards causation — W. S. Anglin
Peacocke on intentional action — John Bishop
Defining intrinsic value — Roderick M. Chisholm
Absolutism v consequentialism: no contest — Don Locke
A utilitarian paradox — Frederick Kroon

BASIL BLACKWELL . ALFRED STREET . OXFORD

MIND

A QUARTERLY REVIEW
OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XC No 358 April 1981

Articles
- **Minimal Rationality:** C. CHERNIAK
- **Against Meaning-Nominalism:** B. C. THURSTON
- **Autonomy of the Will:** N. FLEMING
- **Moral Rules as Expressive Symbols:** J. R. CAMERON
- **Aristotle on the Good for Man:** M. V. WEDIN

Discussions
- **Plato's 'Third Man' Arguments:** F. R. PICKERING
- **Unger's Neo-classical Scepticism:** M. P. SMITH
- **Ockham's Rejection of Ampliation:** G. PRIEST and S. READ
- **Equality: On Sen's Weak Equity Axiom:** J. GRIFFIN
- **Compatibilism: A Reply to Shaw:** R. FOLEY
- **Promises, Promises:** A. D. WOOZLEY

Critical Notice
- **T. Nagel, *Mortal Questions*:** J. GLOVER

Book Reviews

Books Received

Orders to:
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

Annual subscription:
Institutions: £10.00 or $22.50 (Canada $26.00)
Individuals: £8.00 or $18.00 (Canada $21.00)
Single issues: £3.00 or $6.75 (Canada $7.75)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

JOURNAL FOR THE THEORY OF SOCIAL BEHAVIOUR

Edited by P.F. SECORD, (College of Education, University of Houston) and JANET WEST.
Associate Editors: P. MANICAS (Department of Philosophy, Queens College, CUNY)
and C. SMITH (Department of Sociology, Queens College, CUNY).

Volume 11, No. 1 **March 1981**

GERALD L. PETERSON: Historical Self-understanding in the Social Sciences: The Use of Thomas Kuhn in Psychology

JOHN G. ADAIR and BARRY SPINNER: Subjects' Access to Cognitive Processes: Demand Characteristics and Verbal Report

JOSEPH MARGOLIS: Prospects for an Extensionalist Psychology of Action

WILIAM G. NOBLE: Gibsonian Theory and the Pragmatist Perspective

Discussions

EDWARD S. REED and REBECCA K. JONES: Is Perception Blind?
A Reply to Heil

GIFFORD WEARY and JOHN H. HARVEY: Evaluation in Attribution Processes

JOHN P. SABINI and MAURY SILVER: Evaluations in Commonsense Thought:
A Reply to Weary and Harvey

Annual subscriptions: Institutions £20.50 (inland) £25.25 (overseas) $52.50 (USA) $60.00 (Canada); Individuals £10.50 (inland) £13.00 (overseas) $30.00 (USA) $34.50 (Canada)

Orders to: JOURNALS DEPARTMENT, **BASIL BLACKWELL PUBLISHER.,**
108 Cowley Road, Oxford OX4 1JF, England.

"IMPRIME PERMIS NUMERO RT6"

PORT PAYÉ
ROTTERDAM

IF UNDELIVERED PLEASE RETURN TO:-

BASIL BLACKWELL PUBLISHER
c/o MERCURY AIRFREIGHT
INTERNATIONAL LTD
P O BOX 83
ROTTERDAM 3000 AB
HOLLAND

PLEASE QUOTE YOUR
SUBSCRIBER NUMBER
ON ALL CORRESPONDENCE

ANALYSIS 41:2 APRIL 1981

BEGGING THE QUESTION 1981

By RICHARD ROBINSON

THERE is an interesting occurrence of the concept of begging the question in ANALYSIS 39.1, January 1979, p. 24. Burch there writes:

> I would like here to defend Leibnizianism by showing that Plantinga's argument against it fails. The reason it fails is simply that it begs the question: it assumes the existence of a certain possible world whose existence the Leibnizian would deny.

If begging the question were assuming a premiss which somebody might deny, *all* arguments would beg the question.

An argument can well be logically valid in spite of assuming a premiss which somebody might deny. Whether my premisses entail my conclusion is a question independent of the question who denies or accepts my premisses.

Burch writes of an argument *failing*. When does an argument *fail*? I see two cases. An argument *fails* if:

(1) it is logically invalid, or
(2) it does not convince the person to whom it is addressed.

These two ways are independent of each other. An invalid argument may or may not convince the person to whom it is addressed; and an argument which fails to convince a certain person may or may not be a valid argument. Here is an instance of the second kind of failure which is not also an instance of the first kind:

Not not-p, therefore p.

This is a valid argument which does not convince anyone who believes not-p.

To use this second sense of 'fails' can be a little harsh. It seems a little harsh of me if I say that your argument 'fails' because it begs the question because it does not convince somebody because he does not accept your premiss. I had better omit the word 'fails' and the phrase 'begs the question', and say only that you do not convince him because he does not accept your premiss.

Oriel College, Oxford

© RICHARD ROBINSON 1981

WHAT IS WRONG WITH DISJUNCTIVE SYLLOGISM?

By Stephen Read

WE need, and not just for the purpose of this paper, to distinguish between:

(i) if A and $\ulcorner$not A or B$\urcorner$ are true, then so is B,
(ii) if A and $\ulcorner \sim A + B \urcorner$ are true, then so is B,
(iii) if A and $\ulcorner \sim A \vee B \urcorner$ are true, then so is B.

I will call these inference schemata respectively DS_{or}, DS_{+}, and DS_{v}. Each has a claim to be called 'Disjunctive Syllogism'. DS_{or} is the intuitive principle; DS_{+} and DS_{v} formalize it using two different logical accounts of 'or'. '$\vee$' is the usual truth-functional disjunction, for which Addition holds (i.e., if B is true then so is $\ulcorner A \vee B \urcorner$). '$+$' is an intensional disjunction. It is often defined by: $A + B =_{df} \sim A \rightarrow B$, where '$\rightarrow$' is an (again intensional) implicational connective. For it, Addition fails. More will be said about it below.

We can also distinguish these three inferential principles:

(iv) if A and $\ulcorner$if A then B$\urcorner$ are true, then so is B,
(v) if A and $\ulcorner A \rightarrow B \urcorner$ are true, then so is B,
(vi) if A and $\ulcorner A \supset B \urcorner$ are true, then so is B.

I will call these three schemata respectively MP_{if}, $MP_{\rightarrow}$ and $MP_{\supset}$. Each has a claim to be called 'Modus Ponens'. MP_{if} is the intuitive principle; $MP_{\rightarrow}$ and $MP_{\supset}$ formalize it using two different logical accounts of 'if'. '$\supset$' is the usual truth-functional material implication, for which Conditionalization can be vacuous (i.e., if B is true then so is $\ulcorner \sim A \supset B \urcorner$). '$\rightarrow$' is again intensional implication.

'Or' is ambiguous, and (i) can be understood either as (ii) or as (iii). The main purpose of this paper is to show that (ii) is valid while (iii) is invalid. The validity of (ii) will follow from the demonstration of the validity of (v). The invalidity of (vi) will follow from that of (iii). But while (i) is ambiguous, (iv) and 'if' are not ambiguous in the same way. We will see that (iv) is properly represented only by (v), and not by (vi).

II

Recently Bob Meyer has claimed that relevant logic is mistaken in rejecting DS_{v} [1]. He purports to show that it is a valid principle of reasoning, and moreover, one that has a long history of acceptance. In contrast I claim that the rejection of DS_{v} is central to the whole conception of relevant logic.

The central idea of relevant logic is to characterize a connective '$\Rightarrow$' which avoids the paradoxes of material and strict implication, and which expresses the intuitive conception of deducibility, entailment, or valid argument. The paradoxes of material implication show that the truth of $\ulcorner A \supset B \urcorner$ does not suffice for the truth of $\ulcorner$if A then B$\urcorner$, nor for the deducibility of B from A. Moreover, the paradoxes of strict implication show that not even the truth of $\ulcorner A \text{ ⥽ } B \urcorner$, that is, the necessary truth of $\ulcorner A \supset B \urcorner$, suffices for the deducibility of B from A. B is deducible from A if and only if the conditional $\ulcorner$if A then B$\urcorner$ is logically true; and we symbolize this relevantly as

$\ulcorner A \Rightarrow B \urcorner$ is true iff $\ulcorner \Box(A \rightarrow B) \urcorner$ is true,

where '$\Box$' expresses a provability sense of necessity (e.g., S4). That is, we may deduce B from A just when we can prove $\ulcorner A \rightarrow B \urcorner$, i.e. when $\ulcorner A \Rightarrow B \urcorner$ is true. We thus justify this rule of inference:

if A and $\ulcorner A \Rightarrow B \urcorner$ are true, then so is B.

This is $MP_{\Rightarrow}$, Modus Ponens for entailment.

But suppose we cannot prove $\ulcorner A \rightarrow B \urcorner$ *simpliciter*, but can only establish it given some factual assumptions. So we know that $\ulcorner A \rightarrow B \urcorner$ is true, but not necessarily so. Then we cannot conclude that the argument A to B is a valid one, for validity requires the necessary truth of the conditional. Nonetheless we know that, however it might be in general, as things are if A is true then B is too. Hence, given that A is indeed true, we can conclude that B is true. We do not know that B is always true whenever A is, regardless of what other facts obtain. But in the world we are in, B is true when A is, and A is true. So B is true. Thus we have shown that

if A and $\ulcorner A \rightarrow B \urcorner$ are true, then so is B,

which is $MP_{\rightarrow}$.

Thus (v) is a valid principle of inference, central to relevant logic.

III

It follows that (ii) is valid, by the definition of '$+$'. What, however, of (iii) and (vi)?

Despite repeated claims that $DS_{\vee}$ is an intuitively valid rule, it is difficult to find in [1] any argument to this effect. From the argument for γ on p. 87,[1] and in the light of other remarks in the paper, we can reconstruct an argument for $DS_{\vee}$ as follows:

Suppose $\ulcorner A \vee B \urcorner$ is true. Then its truth results either from that of A or from that of B, for '$\vee$' is truth-functional. Hence, if $\ulcorner \sim A \urcorner$ is

---

[1] γ is the principle: if $\vdash A$ and $\vdash A \supset B$ then $\vdash B$. In [2] it is shown that $\ulcorner(\vdash A \,\&\, (\vdash A \supset B) \supset (\vdash B)\urcorner$ is true for all A and B.

> true, that is, A is not true, it must have resulted from the truth of B, and so we can conclude that B is true.

The form of this argument is (i), $\mathrm{DS}_{\mathrm{or}}$: either A is true or B is true; but A is not true; so B is true.

Now 'or' is ambiguous: for ⌜A or B⌝ *can* mean ⌜if not A then B⌝. When a disjunction is used in this way to express a conditional, a modal expression signifying *necessitas consequentiae* is often added. We say for example, 'You must either have posted the letter or burned it' meaning 'If you have not posted the letter then you (must have) burned it'. When so used, the disjunction ⌜A or B⌝ does not follow from A alone—Addition fails. But the presence of the word 'must' obscures this failure; for the rules which fail—that both ⌜it must be that either A or B⌝ and ⌜either A or it must be that B⌝ follow from A—are not unarguably cases of Addition.

To find a clear case of the failure of Addition we must therefore indicate the inferential sense of the disjunction otherwise than by some modal expression. Here is one way:

> 'So either you posted the letter or you burned it.'
> 'You mean you really think that if I didn't post it I burned it?'
> 'That's right.'
> 'Why do you think that?'
> 'Because you posted it.'
> 'But that doesn't follow. Just because I posted the letter, you cannot licitly infer that either I posted it or I burned it—nothing follows about what happened if I didn't post it.'

The last speaker is right: Addition cannot be used to justify the assertion of a disjunction having inferential force. Yet sometimes we do use 'or' in such a way that Addition is valid. 'You qualify for a grant if either you are over 65 or you earn less than £2000 a year. You are over 65. So you qualify.' To satisfy a disjunctive condition it suffices to satisfy one disjunct.

On the first interpretation of 'or', when ⌜A or B⌝ means ⌜if not A then B⌝, (i) is equivalent to (iv). Since '$\rightarrow$' is the formal account of 'if . . . then', the validity of (iv) follows from that of (v). Hence, with 'or' so taken, (i) is valid.

But if 'or' is taken in such a way that Addition is valid, then the disjunction is not equivalent to a conditional, it lacks inferential force, and $\mathrm{DS}_{\mathrm{or}}$ is invalid. This is the moral of the Lewis argument, from A and ⌜not A⌝ to B (let us call this move *consequentia horribilis*). Indeed, if Meyer's defence of $\mathrm{DS}_{\vee}$ were sound, so would be a similar defence of *horribilis*. Meyer notes that $\ulcorner A \mathbin{\&} ({\sim}A \vee B) \rightarrow B\urcorner$ is not a relevant thesis, while $\ulcorner A \mathbin{\&} ({\sim}A \vee B) \supset B\urcorner$ is. The latter, he contends, is sufficient for the deduction of B from $\ulcorner A \,\&\, (\sim A \vee B)\urcorner$ (see [1], p. 88). But $\ulcorner (A \,\&\, \sim A) \supset B\urcorner$ is also a relevant thesis; it would then follow that $\ulcorner A \,\&\, \sim A\urcorner$ entails B.

In Meyer's defence, given above, of DS_v, the 'or' used in the application of DS_{or} is extensional. The reason is what he himself calls the Principle of Parity: essentially that in the truth-condition for 'ξ', '$\ulcorner A \,\xi\, B\urcorner$ is true iff A is true ξ B is true', 'ξ' must not equivocate. So when he infers from the truth of $\ulcorner A \vee B\urcorner$ that A is true or B is true, 'or' here must be '$\vee$'; and so does not suffice for DS. $DS_v(=MP_{\supset})$ is invalid.

What I am arguing is this: if as Meyer claims, the logical truth of

$$((\vdash A) \,\&\, (\vdash A \supset B)) \supset (\vdash B)$$

were sufficient for the deduction of '$\vdash B$' from '$\vdash A$' and '$\vdash A \supset B$', then, by the Principle of Parity, the logical truth of $\ulcorner A \supset B\urcorner$ would suffice for the validity of the argument from A to B, and so one would be committed to classical logic. All classical tautologies are provable in relevant logic.

The danger is an ever present one for the relevant logician. For the classical logician purports to tell him that, by truth-tables, one formula is true when some other is. For example, that B is true whenever $\ulcorner A \,\&\, \sim A\urcorner$ is. But to accept that the truth-tables warrant this claim is to accept the classical account of validity. What the truth-tables tell us is that $\ulcorner A \,\&\, \sim A \supset B\urcorner$ is (logically) true. And that is only a necessary, not a sufficient, condition for validity.

IV

So (i) equivocates between (ii) and (iii), where (ii), being equivalent to (iv), is valid, and (iii), equivalent to (vi), is not. But what actually in itself is wrong with (vi)? Its consequences are wrong, in that it leads to *horribilis*; but what mistake does it make?

The mistake is that $MP_{\supset}$, *horribilis*, and other classical inferences such as disjunctive syllogism are not truth-preserving. Classical validity is often characterized as the logic of truth preservation. If this is accepted then anyone who rejects such inferences has to suggest a criterion for validity stronger than mere truth-preservation. But classical validity is not truth-preserving. For example, it is not true that if $\ulcorner A \,\&\, (\sim A \vee B)\urcorner$ is true then B is true, for any sentences A and B, whatever the classical logician may say about his truth-tables. Certainly $\ulcorner A \,\&\, (\sim A \vee B) \supset B\urcorner$ is a logical truth—it is true whatever A and B may be. But to infer from this that 'if $\ulcorner A \,\&\, (\sim A \vee B)\urcorner$ is true then so is B' is logically true is to suppose 'if' is truth-functional. Given the further (correct) principle that any argument is valid if the corresponding conditional is logically true, classical acceptance of DS_v immediately ensues.

So for validity one must require real truth-preservation, not the classical Ersatz of strict implication. An argument is valid if and only if the conclusion is true whenever the premises are true. DS_v and $MP_\supset$ fail this criterion.

DS_{or} has a long history; for, properly understood, it is valid. But the terms 'Disjunctive Syllogism' and 'Modus Ponens' have for so long been taken for the invalid classical rules DS_v and $MP_\supset$ that the relevant logician cannot but open his case by classing these modes of inference as invalid. He will betray his *modus vivendi* if he accepts as valid their extensional interpretations. Inference is intensional.

University of St. Andrews © STEPHEN READ 1981

REFERENCES

[1] R. K. Meyer, 'Why I am not a Relevantist', Australian National University Research School of Social Sciences, Logic Group Research Paper No. 1 (1978).
[2] A. R. Anderson and N. D. Belnap, *Entailment*, Vol. 1 (Princeton, 1975).

KRIPKE'S SPEAKER'S REFERENCE

By ROD BERTOLET

IN the course of arguing that Donnellan's distinction between referential and attributive uses of definite descriptions is not semantic, Kripke suggests that semantic reference and speaker's reference are 'special cases' of the Gricean distinction between what a speaker's words mean (on an occasion) and what he means in saying those words (on that occasion). I believe that Kripke's suggestion about speaker reference is wrongheaded, and the purpose of this note is to explain why. (See Saul A. Kripke, 'Speaker's Reference and Semantic Reference', in P. French, T. Uehling and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language* (Minnesota, 1979), pp. 6–27. My discussion will be limited to the connection Kripke sees between speaker's reference and Grice's work.)

The example Kripke offers to indicate Grice's distinction is actually a case of what Grice calls conversational implicature; one burglar says to another, 'The cops are around the corner', where what *his words* meant is that the police are around the corner, whereas what *he* meant in saying those words might have been, 'We can't wait around collecting any more loot: Let's split!' (p. 13). (Actually, according to Grice, the burglar will have meant both of these, but let that pass.)

Though he wants his suggestion about speaker reference to apply to descriptions and presumably other singular terms as well, Kripke couches his discussion in terms of the use of a name:

> Two people see Smith in the distance and mistake him for Jones. They have a brief colloquy: "What is Jones doing?" "Raking the leaves." "Jones", in the common language of both, is a name of Jones; it *never* names Smith. Yet, in some sense, on this occasion, clearly both participants in the dialogue have referred to Smith, and the second participant has said something true about the man he referred to if and only if Smith was raking the leaves (whether or not Jones was). How can we account for this? Suppose a speaker takes it that a certain object *a* fulfills the conditions for being the semantic referent of a designator, "*d*". Then, wishing to say something about *a*, he uses "*d*" to speak about *a*; say, he says "$\phi(d)$". Then, he said, of *a*, on that occasion, that it ϕ'd; in the appropriate Gricean sense . . . he *meant* that *a* ϕ'd. This is true even if *a* is not really the semantic referent of "*d*". If it is not, then *that a* ϕ*'s* is included in what he meant (on that occasion), but not in the meaning of his words (on that occasion) (pp. 14–15).

This is followed by a tentative definition of the speaker's referent whose relation to these remarks is less than transparent; I am content to leave this as a problem for Kripke, leaving open the possibility that the definition, whatever its other merits, does not inherit the flaws of these remarks. But I do wish to argue that these remarks are flawed.

The difficulty begins to emerge when we reflect on how exactly we are to specify what the speaker meant. Kripke claims that he, as opposed to his words, meant that *a* ϕ's: but apparently *a* is an individual taken independently of any way in which that individual is specified. (So, *that a* ϕ*'s* is perhaps best thought of as a singular proposition of the sort in vogue in some circles, representable by an ordered pair consisting of an individual and a propositional function.) But in *Grice's* sense of speaker meaning, what the speaker means has to be specified in some fuller way (by a more old-fashioned Fregean proposition), and the problem is that there is no obvious candidate for this role. *That Jones is raking the leaves* is out; this is what the speaker's *words* meant, and will not yield a distinction between this notion and what he meant. *That Smith is raking the leaves* is out for another reason: the speaker is unaware that the leaf-raker is Smith, so he cannot have *meant*, in the sense Grice intends, that Smith is raking the leaves.

Certainly he cannot have meant this on Grice's *analysis* of speaker meaning, which postulates a series of nested intentions which include the intention that the hearer think, or think that the speaker thinks, that Smith is raking the leaves (see 'Utterer's Meaning and Intentions', *Philosophical Review* 78, 1969, for the view I have just oversimplified tremendously). Kripke's speaker, who is unaware of the leaf-raker's being Smith, is in no position to have such an intention. Now Kripke need not be wedded to Grice's analysis (and, he claims in a footnote to 'loosely' follow Grice, and wonders on p. 14 whether one of his examples refutes Grice's analysis). Still, in the intuitive sense of what is meant to which Grice appeals, and which Kripke exploits in the burglar example, the speaker must be *aware* of what he means, and this is not the case here: he is ignorant of the leaf-raker's being Smith. Nor is there any other clear candidate for what the speaker meant, in the sense of 'meant' on which Grice is relying. (This is perhaps more plainly so in Donnellan's original man with the martini example, where the speaker has no description or other linguistic device which correctly and uniquely applies to the referent—thus there is no sentence/proposition such that we can say the speaker meant that . . . , where the referent is correctly identified by what replaces the ellipsis.)

There is, then, no easy assimilation of the example Kripke discusses to Grice's distinction between what a speaker's words mean and what he means by them or in saying them. Of course it would be nice if there were an assimilation; Grice's distinction is motivated by examples quite different from the ones Kripke and Donnellan are concerned with, and is accompanied by an apparatus designed to explain just how what the speaker meant is communicated. So it would certainly be handy to plug the notion of speaker reference into this independently motivated machinery—however, Kripke's attempt to make the connection seems untenable.

I do not want to deny that there is some intuitive appeal to the claim that the speaker meant that a ϕ'd (independently of any way of specifying a); I do deny that this can be assimilated in any easy way to Grice's notion of speaker meaning, where the speaker means *that* something is the case. (I am thus not as hostile to singular propositions as some, or as my previous remarks might be taken to suggest.) It is evident from the literature that (at least) some philosophers find such a specification of what someone *means* rather too thin, roughly for Frege's reasons. To combat their intuitions and arguments, to explicate and defend the use of singular propositions, will require a good deal more work: my point has been that Kripke is wrong to think that Grice can be called upon to abet the endeavour.

Purdue University © ROD BERTOLET 1981

THE DISCREET CHARM OF COUNTERPART THEORY

By GRAEME HUNTER *and* WILLIAM SEAGER

I

MANY philosophers believe David Lewis's Counterpart Theory suffers from insoluble foundational difficulties (the most telling criticisms and those on which we will focus are due to Allen Hazen's 'Counterpart-theoretic Semantics for Modal Logic', *Journal of Philosophy* 76, 1979, pp. 319–38). According to them, the truth value of certain modal sentences is not preserved in Counterpart Theory. In our view however this difficulty arises from the uncritical use of the translation scheme originally proposed by Lewis (in 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 65, 1968, pp. 113–26) and taken by many to be integral to the theory rather than heuristic. Counterpart Theory unfettered by this scheme is not only proof against their objections but can provide elegant solutions to certain traditional problems of modality and metaphysics.

Consider the infamous sentence (diffidently presented even by Lewis):

(1) Everything actual necessarily exists.

The standard QML translation of (1) is:

$$(2)\ (x)L(Ey)(x = y).$$

Now the Counterpart Theoretic translation of this according to Lewis's scheme is:

$$(3)\ (x)(Ix\alpha \rightarrow (x1)(y1)(Ix1y1\ \&\ Cx1x \rightarrow (Ex2)(Ix2y1\ \&\ x2 = x1))$$

(where 'Ixy' means 'x is in possible world y', 'Cxy' means 'x is a counterpart of y' and 'α' denotes the actual world). Unfortunately, (3) is true and for that matter necessarily so. This can easily be seen since it amounts to no more than the obvious truth that any counterpart of an actual object is identical to something. However, (1) is certainly false for surely not everything had to exist.

So if Counterpart Theory commits us to (3) then Counterpart Theory is wrong. But it doesn't, only the translation scheme does. Do we then have a new translation scheme? No, it is no more necessary to have one to capture the sense of modal idiom than it is to capture the sense of quantificational idiom. Still, some translations are clearly better than others and some are plainly wrong. Appeal to truth value is the single most important criterion of correct translation.

Let's consider some alternatives to (3). Counterpart Theory suggests that an object has a property necessarily just in case all its counterparts have it, and it exists just in case it is in this (the actual) world. So it might seem that to exist necessarily is for an object to have all its counterparts in the actual world. This suggests that we replace (3) by:

(4) $(x)(Ix\alpha \rightarrow (x1)(Cx1x \rightarrow Ix1\alpha))$.

This proposal admirably illustrates the difficulties of translation. (4) *is* false, but it says more than (1). Indeed, on Lewis's original formulation of Counterpart Theory (4) would commit us not only to the necessity of the existence of actual objects but to the necessity of all their properties as well, since it is a postulate (P5, op. cit., p. 114) that nothing has any counterpart other than itself in its world. But even if this postulate were weakened, surely nothing has enough actual counterparts to guarantee the contingency of all the properties which we think should be contingent. And surely the necessity of our existence should not commit us to the necessity of all or most of our properties. The desideratum is then a formula which, like (4), is false but which, unlike (4), captures the notion of necessary existence without committing us to the necessity of other properties.

An alternative and perhaps more natural way to conceive of an object's necessary existence is for the object to exist in all possible worlds. The Counterpart Theoretic *pendant* of this is:

(5) $(x)(y)(Ix\alpha \rightarrow (Ex1)(Ix1y \,\&\, Cx1x))$.

That is, any actual object has counterparts in every possible world. (5) is false for the same reasons (1) is. And it also makes no untoward commitments regarding the properties of actual objects. So this alleged counterexample, while devastating to the translation scheme which Lewis gave us, leaves the underlying theory unscathed.

II

Another objection of Hazen's (op. cit., p. 326) is that Counterpart Theory fails to validate the inference from:

(6) LRab

to:

(7) L(Ex)Rax.

As might be expected it also fails to validate the inference from (6) to:

(8) L(Ex)Rxb.

This is so because Counterpart Theory is expressively richer than QML. In Counterpart Theory (6) is quadruply ambiguous. There are three *de re* readings and one *de dicto* reading and not every reading validates these inferences.

As a case of (6), let a be the child of b and suppose that this is a necessary relation. But necessary for whom? For a, b or both? If for a then the inference from (6) to (7) is reasonable, but that from (6) to (8) is less so. For surely even your being essentially your father's child does not foredoom your father to incelibacy. While it is most natural to interpret sentences like (6) as giving an essential property of a, this need not be the case. Thus the Counterpart Theoretic version of the normal case would be:

(9) $(x)(y)(Cxa \ \& \ Ixy. \rightarrow (Ex1)(Cx1b \ \& \ Ix1y \ \& \ Rxx1)).$

That is, in any world where there is a counterpart of a there is also a counterpart of b to which a is R-related. Contrary to Hazen's claim, on the normal interpretation of (6) the Counterpart Theoretic equivalent of (7) does follow:

(10) $(x)(y)(Cxa \ \& \ Ixy. \rightarrow (Ex1)(Ix1y \ \& \ Rxx1)).$

But of course the Counterpart Theoretic equivalent of (8):

(11) $(x)(y)(Cxb \ \& \ Ixy. \rightarrow (Ex1)(Ix1y \ \& \ Rx1x))$

will not follow from (9) since (9) imposes no necessity on counterparts of b to exist only in worlds where there are also counterparts of a. The unusual case of (6) being taken to give an essential property of b:

(12) $(x)(y)(Cxb \ \& \ Ixy. \rightarrow (Ex1)(Cx1a \ \& \ Ix1y \ \& \ Rx1x))$

is exactly analogous, supporting (11) instead of (10). Lastly, if (6) is taken to state a jointly essential property of a and b, then the Counterpart Theoretic equivalent of (6) is just the conjunction of (9) and (12). So in any case we can validate a correct inference and invalidate an incorrect one merely through proper translation with no need to resort to 'complex predicates' or otherwise to complicate the relational structure we begin with.

But there is an obvious counter-argument which seems to maintain the validity of the inference to either (7) or (8), namely the following QML-valid sequence:

LRab
Rab
(Ex)Rax
L(Ex)Rax.

(The actual form of this proof will depend on the details of the logical system used.) But the point is that only on a *de dicto* reading of 'LRab' will the inference go through. The *de dicto* reading guarantees that 'Rab' will be true in every possible world, hence in ours, thus validating the existential generalization. The *de re* reading says only that a bears R to b essentially and does not demand that a or b exist in our world. We agree that one can make this inference in QML, but this is hardly surprising given that a *de dicto* reading of 'LRab' is absurdly strong for most values of 'R'.

III

Another merit of Counterpart Theory is that it forces us to see the distinction between essential existence and essential properties. The attentive reader may have noted that the instantiation of (2):

(13) $L(Ex)(a = x)$

is of the form of (7). One might then suspect that the Counterpart Theoretic translation of (13) ought to be:

(14) $(x)(y)(Cxa \,\&\, Ixy. \rightarrow (Ex1)(Ix1y \,\&\, x = x1))$.

But once again this illustrates that rote translation doesn't work in general, for (14) makes only the fatuous claim that every counterpart of a is identical to something, which is somewhat less than necessary existence. We have, in fact, already shown that the necessary existence of every *actual* object is captured by (5) above. So by parity of reasoning the Counterpart Theoretic equivalent of (13) should be:

(15) $(y)(Ex)(Cxa \,\&\, Ixy)$.

A moment's attention to (10) and (15) will reveal that they are as similar as they can be without the collapse of the distinction between properties and existence which led to the trivial (14).

Thus the present section underlines the Kantian distinction between property and existence, with obvious ramifications for ontological arguments. The second section shows the relational structures involved in standard quantificational theory can be maintained without prejudice to the traditional intuition that relata are not always on an equal footing, an intuition which led to the logic of terms. The first section showed the expressive adequacy of Counterpart Theory but went only a short way towards revealing its richness. As in the case of quantificational theory, time spent delving beneath the surface grammar with the tools of Counterpart Theory will reveal fundamental patterns of our reasoning.

University of Toronto GRAEME HUNTER *and* WILLIAM SEAGER 1981

ACTIONS, REASONS, AND HUMEAN CAUSES

By PETER H. HESS

IN this paper I shall dispel an illusion. Numbered amongst the victims of this illusion are many of those who subscribe to the time-honoured view according to which the causes of an action include, among other things, the agent's beliefs together with his desires, wants, intentions, etc. The reason why they labour under an illusion is because they think that they can find support for their position in the writings of Donald Davidson. I shall show not only that Davidson's views, as expressed in two of his articles, do not corroborate their position, but that his remarks, once they are carefully analysed, turn out to be so jejune as to be acceptable even to those who think that the traditional thesis ought to be completely abandoned.

The first of the two Davidson articles that I wish to consider, [1], consists of an attempt to argue in favour of the following two claims:

> C1: R is a primary reason why an agent performed the action A under description d only if R consists of a pro attitude[1] of the agent toward actions with a certain property, and a belief of the agent that A, under the description d, has that property ([1], p. 687).
>
> C2: A primary reason for an action is its cause ([1], p. 693).

Since I do not have any quarrel with C1, I straightaway turn to consider C2. In his attempt to set the scene for his defence of C2, Davidson offers the following considerations: Since people frequently perform actions which they have a reason to perform even though that reason is not why they acted in that way, we need a scheme that gives 'an account of the "mysterious connection" between reasons and actions' and which applies in those cases where an agent acts *because* of his reasons ([1], p. 693). C2 alone provides us with such a scheme and allows us 'to turn the first "and" to "because" in "He exercised *and* he wanted to reduce and thought exercise would do it" ' ([1], p. 693). I shall show that C2, if properly interpreted, proves unable to carry the burden Davidson shoulders it with. It, consequently, cannot be the *only* scheme that explains the 'mysterious connection' between reasons and actions.

We discover how weak C2 really is when we turn to Davidson's discussion of the Humean notion of causation.[2] Davidson agrees with

---

[1] Pro attitudes, according to Davidson, include 'desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent towards actions of a certain kind' ([1], p. 686).

[2] Part of his reply to objection C; [1], pp. 696–99.

Hume's contention that every true singular causal claim entails a corresponding true causal law. Yet he also admits that there aren't any laws connecting primary reasons with actions. In order to show that these views are not incompatible with C2, he argues that Hume's claim is ambiguous:

> It may be that 'A caused B' entails some particular law involving the predicates used in the descriptions 'A' and 'B', or it might mean that 'A caused B' entails that there exists a causal law instantiated by some true description of A and B ([1], p. 698).

Davidson opts in favour of the second of these two interpretations. 'Only the second version of Hume's doctrine,' he says, 'can be made to fit with most causal explanations; it suits rationalizations [i.e., explanations referring to primary reasons] equally well' ([1], p. 698). He illustrates the error involved in the assumption that 'singular causal statements necessarily indicate, by the concepts they employ, the concepts that will occur in the entailed law' ([1], p. 698) with the following example:

> Suppose a hurricane, which is reported on page 5 of Tuesday's *Times*, causes a catastrophe, which is reported on page 13 of Wednesday's *Tribune*. Then the event reported on page 5 of Tuesday's *Times* caused the event reported on page 13 of Wednesday's *Tribune* ([1], p. 698).

It would clearly be erroneous here to look for a law connecting events of this kind. Hence Davidson concludes that 'The laws whose existence is required if reasons are causes of actions do not, we may be sure, deal in the concepts in which rationalizations must deal' ([1], p. 699). He suggests that these laws may even be formulated in terms of classifications that are neurological, chemical, or physical ([1], p. 699).

Before we determine how the matters just discussed affect the meaning of C2, let us look in some more detail at Davidson's example. We can agree, I suppose, that the event reported in the *Times*, apart from having the property of having been reported in that newspaper, also had certain other features the presence of which enabled it to cause a catastrophe. I am sure we can also agree that the hurricane would have caused a catastrophe, even if it had not been reported in the *Times*. The reason for this is two-fold: First, there have been and will be lots of hurricanes like the one described in the *Times* which have caused and will cause catastrophes like the one reported in the *Tribune*, even though neither they nor their consequences have been or will be described in either paper. Thus not every pair of the properties of the events in question can be incorporated into a true generalization. Furthermore, even if this could be done, not all of the resulting true generalizations would be law-like, i.e., not all of them would support counterfactual and subjunctive statements. In fact, the properties 'being reported in the

Times' and 'being reported in the *Tribune*' seem to figure in no law-like generalization under which the two events (i.e., the hurricane and the catastrophe) can be subsumed. And this shows that the presence or absence of these properties in no way affects the causal efficacy of the events whose properties they are. If we keep this in mind we will not be misled by such claims as 'The event reported in the *Times* causes a catastrophe'. We will not assume that one can diminish the havoc-wreaking potential of a hurricane by keeping quiet about it.

What, then, does all of this tell us about C2? It tells us that we ought to deal with the claim 'Primary reasons are the causes of actions' with as much perspicuity as we previously employed in dealing with the claim 'Subject matter of *Times* report causes catastrophe'. Davidson clearly intends C2 to assert that actions have causes and that these causes include the presence of primary reasons. His remarks, furthermore, suggest that he considers C2 to be perfectly compatible with the claim that the causal efficacy of whatever causes an action is in no way affected by the presence or absence of primary reasons.

Advocates of what Davidson calls 'the ancient—and common-sense—position that rationalization is a species of ordinary causal explanation' ([1], p. 685) should not derive any comfort from what he offers them in the form of C2. They will agree that actions have causes (but, pace Melden [2], so does just about everyone else) and they will not object to the claim that these causes include the presence of primary reasons. But they certainly will not accept the view that the causal efficacy of whatever causes an action is in no way affected by the presence or absence of primary reasons. In view of this they will not even be able to claim that C2 gives at least an acceptable account of the 'mysterious connection' between reasons and actions, for C2 does not. It merely reduces primary reasons to the status of epiphenomena. This being the case, it is time for the denouement of the illusion that Davidson speaks on behalf of the traditional view.

Lest it be thought that this way of interpreting C2 depends too heavily on the use it made of one particular illustration, I shall now show that it is compatible with what Davidson says in another article, [3], in which he also discusses the causal efficacy of mental events. I shall also try to explain more clearly what I meant when I argued in the above that in Davidson's scheme of things primary reasons are reduced to the status of epiphenomena. In that second article Davidson endeavours to show that the following three principles are not inconsistent with one another:

(1) The Principle of Causal Interaction, according to which 'at least some mental events interact causally with physical events' ([3], p. 80).

(2) The Principle of the Nomological Character of Causality, according to which 'events related as cause and effect fall under strict deterministic laws' ([3], p. 81).

(3) The Principle of the Anomalism of the Mental, according to which 'there are no strict deterministic laws on the basis of which mental events can be predicted and explained' ([3], p. 81).

The reason why Davidson thinks that each one of these principles is true, and thus the reason why he does not think that their conjunction will lead to a contradiction, is that he advocates a version of the Psycho-Physical Identity Theory which he dubs 'Anomalous Monism'. This theory asserts (1) that every mental event is identical with some physical event and (2) that there are no correlating laws which relate the mental to the physical. The reason why, according to Davidson, there are no such laws is that even though there may be true general statements linking the mental with the physical, none of these statements will be law-like.

I will now show that by accepting anomalous monism one reduces primary reasons (or, more precisely, the mental properties an event has by virtue of the fact that it involves the presence of a primary reason) to the status of epiphenomena.

I begin with the following account of what I mean by 'epiphenomenal':

> A property P is epiphenomenal with respect to the relationship between an event C and its effect E iff
> (i) P is a property of C;
> (ii) It is not the case that C would not have caused E had it not had property P.[3]

If the theory of anomalous monism is true, then every mental property is epiphenomenal with respect to every mental event and its physical consequences. If there were even one mental property that is not epiphenomenal in that sense, then there would be at least one psychophysical law. But, according to Davidson, there aren't any such laws. Thus it follows that mental events possess the same causal potency that newsworthy events (such as hurricanes that have been described in the press) possess. Just as the latter do not lead to catastrophes because they have been reported in the press, so the former do not cause physical events because they have mental properties. Neither, of course, do they cause physical events because they involve the presence of primary reasons. All of this may, in fact, be true. The arguments that Davidson marshals in support of the three principles that were previously mentioned strongly suggest that it is true. But it is hardly a state of affairs that will comfort those who subscribe to the traditional view according to which mental events cause physical events. The traditional view, surely, includes the assumption that there are at least some mental properties which are not epiphenomenal with respect to some mental events and their physical consequences.

---

[3] It is important to distinguish between this condition and one which states that 'It is not the case that E would not have happened, had P not been a property of C'. To see the difference, consider an action E (e.g., 'John turns on a light-switch') and an event C the occurrence of which was causally sufficient for the occurrence of E and which, furthermore, included the presence of a belief P (e.g., 'John believed that he could illuminate the room by turning on the light-switch'). An anomalous monist could not remain true to his belief that there are no law-like connections between the mental and the physical and, at the same time, claim that if C had not included the belief in question, C would not have caused E. He may, however, still be able to assert that if the agent had not entertained the belief which he did entertain, he would not have acted as he did. After all, there may be subjunctive and counterfactual conditionals which are true and contingent (i.e., unlike 'Had the number of John's children been even, he would not have had three children') and which, nevertheless, are not subject to a causal interpretation.

This completes my argument. Let me make the following, concluding remarks:

I have endeavoured to explain why advocates of the traditional view concerning the causal efficacy of reasons should not turn to Davidson for support. I leave it to others to decide how important it was to make that point. But there is another issue which my paper, at least implicitly, deals with and which is of more far-reaching significance. Davidson's theory of Anomalous Monism attempts to show that the traditional view concerning the causal efficacy of primary reasons is compatible with the thesis which states that there are no strict psychophysical laws. Davidson's arguments in favour of the claim that the mental is nomologically irreducible are, as I said, most impressive. If, in view of this, one still wants to claim that reasons are causes, one must alter the traditional view so as to allow for the epiphenomenal character of the mental. To do this, and this has been the thesis of my paper, is to abandon the traditional view.

Victoria College, University of Toronto © PETER H. HESS 1981

REFERENCES

[1] Donald Davidson, 'Actions, Reasons, and Causes', *The Journal of Philosophy* LX, No. 23 (Nov. 7, 1963), pp. 685–700.
[2] A. I. Melden, *Free Action*, New York, 1961.
[3] Donald Davidson, 'Mental Events', in L. Forster & J. W. Swanson (eds.), *Experience and Theory*, New Haven, 1971.

ON MACKIE'S NEW ACCOUNT OF CAUSAL PRIORITY

By DOUGLAS EHRING

J. L. MACKIE in 'Mind, Brain, and Causation' (*Midwest Studies in Philosophy* 4, 1979, pp. 19–29) offers a new account of causal asymmetry which differs from his earlier one in *The Cement of the Universe* (Oxford, 1974). After briefly stating his new view I will present four examples which run counter to his approach.

Mackie's general statement of his new position on causal priority runs as follows:

> . . . if on a particular occasion A's doing X is causally related to B's doing Y, and if they had not been so related but things had otherwise been as far as possible as they were, A would still have been doing X but B would (or might) not have been doing Y, then A's doing X is conditionally and causally prior to B's doing Y. And this holds whether A's doing X is necessary for B's doing Y, or sufficient, or both (op. cit., p. 24).

Mackie illustrates this view with an example of two causally related items, the rotation of a drive shaft and the rotation of the rear wheels of a car. When the engine is moving the car along the road, the rotation of the wheels is causally dependent on that of the drive shaft. What does this involve? '. . . this means that if the connection between the two rotations had not been there—say, if the differential had failed or had been missing—but things had otherwise been as far as possible the same, the drive shaft would have been rotating but not the wheels' (ib., p. 24). As Mackie puts it, the directed conditionality is formulated in terms of a counterfactual conditional of which the contrary-to-fact antecedent is that the causal relation between the relevant items is absent. The antecedent also includes the condition that the failure of the causal relation is the result of a 'small change in the circumstances' so that the circumstances are as far as possible the same. The consequent of the conditional states that the cause-event occurs but the effect-event fails (or might fail) to occur.

Four cases make it clear that this analysis does not provide an adequate account of the locution 'c causes e':

(1) Suppose that c causes e. A small change in the circumstances is introduced which eliminates the causal connection between c and e. We can further suppose that c's causal efficacy in producing e preempts the causal role of some other event d, which would produce e if c fails to cause e. Thus when the causal connection between c and e is eliminated, d causes e. Mackie's condition has been fulfilled, i.e., the causal relation between c and e has been eliminated, but c and e occur since e has an alternative cause, d. Furthermore, we can suppose that every small

change which will eliminate the causal connection between c and e has the same result of allowing some preempted cause to be effective.

(2) This difficulty may be deepened along the following lines. c causes e. A necessary condition for the causal relation, f, is eliminated. f in turn is a necessary condition for c so that the absence of f in the circumstances brings about the absence of c. At the same time we may suppose that c was a preempting cause of e so that c's failure to occur brings into play d, which causes e. Thus a small change in the circumstances which eliminates the causal relation between c and e has resulted in the absence of c and the occurrence of e. On Mackie's analysis we would be forced to say e causes c contrary to our assumption. Again we may assume that every small change which is sufficient to eliminate the causal relation is sufficient to eliminate c.

(3) Again suppose that f is a necessary condition of the causal relation between c and e, where c causes e. Elimination of f will eliminate the causal relation between c and e. However, we can suppose that f is also a necessary condition for both c and e (for example, f may be the presence of oxygen and c and e two fires). Thus if f fails to occur so will c and e, and Mackie's analysis will not provide a verdict on which is the cause relative to the other. Finally, we may imagine that the only small changes sufficient for the elimination of the causal relation between c and e involve the failure of conditions necessary for c and e.

(4) Mackie claims that his analysis is true whether 'A's doing X is necessary, sufficient, or both' for B's doing Y. Suppose that c is a sufficient but not a necessary cause of e such that e is causally overdetermined by c and d. We then introduce a small change in the circumstances which eliminates any causal relation between c and e. Since e is causally overdetermined e occurs in the absence of this causal link to c. If we further suppose that the small change made also eliminates a necessary condition of c, then we are required to conclude that e causes c contrary to our assumption.

These four counterexamples show that Mackie's analysis does not provide an adequate reconstruction of our concept of causal priority.

Columbia University

© DOUGLAS EHRING 1981

CAUSAL PRIORITY AND THE DIRECTION OF CONDITIONALITY

By J. L. MACKIE

DOUGLAS EHRING's objections (ANALYSIS, this issue) give me a welcome opportunity to explain my approach more clearly. The proposal about causal asymmetry or priority to which he refers was not my second, but my third attempt to deal with this problem (the first was in 'The Direction of Causation', *Philosophical Review* 75, 1966, pp. 441–66), and it may well not be my last. Indeed I indicated how tentative this proposal was in a passage that comes two pages after the ones Ehring quotes:

> What I have sketched here is a conceptual analysis of causal asymmetry based on the direction of conditionality, and a related epistemic analysis. Since I deny that counterfactual conditionals with unfulfilled antecedents (other than those whose antecedents entail their consequents) can be simply and straightforwardly true, I cannot offer it also as a factual analysis, as an answer to the ontological question about causal asymmetry. But such asymmetry in the objects will consist at least in those contrasting regularities, those patterns of resemblances and differences, the detection of which gives rise to and supports those counterfactual judgments. It is a further question whether this account can be developed so as to accommodate my earlier suggestion that the direction of causation is linked with the contrast between fixity and nonfixity, and that an effect is fixed only by way of its cause. Leaving this aside, the account so far given may be enough to throw some light on the causal relations between mental and physical features . . . ('Mind, Brain, and Causation', p. 26).

I should, therefore, be not greatly disturbed if it turned out that there are counterexamples to the formulation that Ehring quotes. My object has been not so much to find a watertight formula as to throw some light on our concept of causal asymmetry, on our methods of detecting it, and ultimately on whatever asymmetry there may be in the processes themselves. The main suggestion in my 1979 article was that we have a certain notion of a direction of conditionality which contributes to our concept of causal direction and that there are typical procedures and observations which confirm such a direction of conditionality in particular cases. Nothing that Ehring has said even tends to show that this was a mistake, so that the most that his alleged counterexamples could show is that I should adjust my formulation to meet them, while retaining the main idea suggested in that article. The 'counterexamples' all involve complications; but surely our first task is to try to see wherein the asymmetry between a cause and its effect lies in a simple, uncomplicated, case, where there are no such obscuring features as overdetermination.

But it is not clear that the 'counterexamples' show even that any reformulation is needed; of the four Ehring gives, three and a half, at any rate, do not. The complications they involve ignore the qualification which was already present in the passage he quotes, '. . . but things had otherwise been as far as possible as they were'. This is equivalent to saying, in the currently popular jargon, that we should consider the nearest possible worlds in which the causal relation fails. In Ehring's (1) we cannot suppose that in the *nearest* possible worlds where c fails to cause e the alternative cause d is brought into play. The same objection applies to (2): there will be nearer possible worlds in which the c–e relation is broken without the removal of a necessary condition for c and without d's being brought into play. In (3) Ehring suggests a concrete instance, where c and e are two fires and the f which is necessary both for the causal relation and for c and e separately is the presence of oxygen; but a nearer possible world in which the connection fails is one in which there is insufficient oxygen just in the region between the two fires, but still enough where each fire is itself. Similarly in (4) we cannot 'further suppose that the small change eliminates a necessary condition of c': there will be a nearer possible world where it does not. Admittedly experiments might yield misleading results, if we failed in practice to alter one feature (the connection) without disturbing others; but that risk does not undermine the proposed conceptual analysis.

That leaves the other half of (4), the case where e is overdetermined by c and d. When I claimed that my analysis 'holds whether A's doing X is necessary for B's doing Y, or sufficient, or both' I was not in fact referring to overdetermination, but to different possible types of causal law (as in *The Cement of the Universe*, pp. 40–43 and 58). It is controversial whether pure cases of overdetermination, where exactly the same effect results from two causes together as would have resulted from either on its own, ever occur. However, let us suppose that we do have such a case. Then indeed we do not have a direction of conditionality as I have formulated it between c and e, though we do have one sort of conditionality in that c is sufficient for e in what I have called (*The Cement of the Universe*, pp. 39–40) the strong, counterfactual, sense of 'sufficient'. But it is doubtful whether our concept of causation allows us to say, in this case, that c causes e. (We can, of course, say that the combination of c and d causes e, but the direction of conditionality as I described it also holds between this combination and e.) If we would say that c causes e here, this must be because we believe (relying on analogy with other cases involving events of the same types) that c would have caused e even if d had not occurred. That is, we may think of there being a direction of causation between c and e where (because of the overdetermining rival cause d) there is not the usual direction of conditionality between c and e, provided that there would have been this direction of conditionality if the effect had not been overdetermined. If we do allow each of two or more overdetermining 'causes' to cause the effect, such an additional clause should be attached to the account of our concept of causal asymmetry which I gave in my 1979 article.

University College, Oxford

© J. L. Mackie 1981

BACKWARDS CAUSATION

By W. S. Anglin

IT is possible to act in such a way that something is made true about the past which would have been false had the act not been performed. For example, if Professor Smith predicts that I shall run off with Maid Marilyn then if I do run off with her I make it true that Smith was right. It is also possible to have it in one's power to act in such a way that, were one now to exercise that power, it would have to have been the case that the past was actually different. For example, if I have been taught how to make a Bloody Mary but Marilyn has already asked for a Screwdriver, then I have it in my power to make a Bloody Mary but, were I to exercise that power, it would have to have been the case that Marilyn did not ask for a Screwdriver. Neither of these two possibilities, of course, implies that I have any real power over the past. What I do or can do may have a bearing on which descriptions are true of the past or which things might have been true in the past but there is no suggestion that a present act could have a real effect on what has already happened. For example, Smith's having been right about me and Marilyn is not a fact that makes any difference to the past itself. The same Smith would have made the same prediction in the same circumstances even if I had not subsequently run off with Marilyn. The purpose of this paper, however, is to defend the logical possibility of a full-blooded backwards causation against two arguments to the contrary.

The first argument against the possibility of backwards causation says that, were it possible, then one could either prevent the cause from happening after the effect had already happened or make the (sufficient) cause happen after the effect had already not happened. Since it is impossible to have an effect without a cause or vice versa, it is impossible to have backwards causation. For example, in *The Cement of the Universe* Mackie entertains the possibility of someone with alleged powers of precognition being able to draw a picture of a pattern which will later on be 'the product of what is in a very strong sense a random device, something which operates indeterministically, by objective chance' (pp. 175–76). Since the origin of the pattern is perfectly random, it cannot be that the picture is a cause of the pattern. Since the picture is a perfect copy of the pattern, it seems there is a causal connection between the two. Thus, it seems, the future pattern is a cause of the present drawing. Mackie, however, says that we could not so interpret the situation because

> on every occasion, after the drawing is made, it is possible that someone or something should intervene so that the corresponding pattern fails to be produced. Consequently, it cannot on any occasion be the pattern that is responsible for the details of the drawing: the precognition hypothesis must be false even for those occasions when the device is not stopped, when the pattern actually is produced and turns out to be just like the drawing (p. 178).

Mackie concludes that if the patterns are really produced in a way that is not somehow determined or fixed in the past, then the resemblance of drawings to patterns would have to be 'fortuitous, however implausible this may seem' (p. 179).

Mackie's argument fails because it presupposes that the past was whatever it was regardless of what occurs in the present or future and this, precisely, is what is at issue. The question we are discussing is whether it is logically possible that a past event be subject to a present event and therefore not fixed or given in any sense in which a future event is not. Let C be a cause and E its effect. Suppose that C is necessary and sufficient for E. In the case in which C occurs before E, if C does not happen then E will not happen but if C does happen then E will happen. Whether or not E occurs is fully explained by saying that C did or did not occur. Similarly, in the case in which C occurs after E, if C does not happen then E did not happen but if C does happen then E did happen. Whether or not E occurs is fully explained by saying that C will or will not occur. Thus if you prevent a past-directed cause C from happening you bring it about that the past effect E did not occur. The reason it did not occur is simply that its cause will not occur. If you make a past-directed cause C happen you bring it about that the past effect E did occur. The reason it did occur is simply that its cause will occur. Thus, whatever you do, you cannot leave E without C or bring about C without E. For example, if in certain circumstances the occurrence of an event E implies that someone prevents its subsequent cause C from occurring, then neither E nor C occurs under those conditions. The answer to the above argument is that although one *could* prevent C, the fact that E has occurred means that one does not. If one was going to prevent C then E would not have happened and for the simple reason that its cause was not going to happen. Put in terms of Mackie's example, this is to say that even if one has the ability to destroy the pattern-producing device, one does not exercise that ability in those cases in which the drawing of the pattern has already been produced, for if one were going to destroy the device the drawing would not have been produced. It would not have been produced because the cause of its production was not going to happen.

Let us consider these ideas in terms of a version of the Newcomb Paradox. Placed before you is a title-deed to a beautiful mansion somewhere in the world and a large sealed envelope. You have the choice of taking just the sealed envelope or else taking both the sealed envelope and the title-deed. This latter option we shall call 'being greedy'. Thanks to backwards causation (and without any possibility of mistake), God has already observed what you will do and has already done the following:

> If you will choose just the sealed envelope, he has put in it a title-deed to a beautiful mansion somewhere in heaven;
> if you will choose both the sealed envelope and the title-deed to the beautiful mansion somewhere in the world, he has left the sealed envelope empty.

Assuming you want the heavenly mansion much more than the earthly mansion, what should you do? It might seem that you ought to be greedy since God has already either put the title-deed to the heavenly mansion in the sealed envelope or he has not. If he has, then you get both mansions, and if he has not, you get the earthly mansion along with a free envelope. To paraphrase Don Locke ('How to Make a Newcomb Choice,' ANALYSIS 38.1, January 1978, p. 18), either there is or is not a title-deed in the sealed envelope, and either way you are better off being greedy. However, as André Gallois has pointed out ('How Not to Make a Newcomb Choice,' ANALYSIS 39.1, January 1979, p. 50), this argument is very similar to the standard argument for fatalism. If we are taking the hypothesis of backwards causation seriously, then the past event of God's putting or not putting the title-deed in the envelope, although it cannot, of course, change, is not fixed and unalterable in any way in which present or future events are not fixed and unalterable: it was whatever it was depending on your future choice. You will decide what God did do. You cannot change the past by your choice but you do affect it. A past event which really depends on a present choice is no more or less fixed or unalterable than a future event which depends on a present choice.

A second argument for being greedy is that an 'intelligent, well-informed and perfect well-wisher' who knew whether the sealed envelope was empty or not would advise you to be greedy. Since what an intelligent, well-informed and perfect well-wisher would advise you to do would be in your best interests then you should indeed be greedy (Don Locke, p. 19). However, if the envelope is full, the well-wisher (being intelligent and well-informed) would know that it is full only because you will reject any advice to be greedy. Thus he would *not* advise you to be greedy because intelligent people do not waste their breath giving advice they know will be rejected. Again, if the envelope is empty, the well-wisher would know that it is empty only because you will foolishly choose to be greedy. Thus he would *not* advise you to be greedy because well-wishers do not contribute to the folly of those whom they wish well. The point is that a well-wisher knowing the contents of the envelope presupposes God's having made his past decision and that in turn presupposes your going to make your future decision. The well-wisher is confronted with a *fait accompli.* It is not temporally but causally 'too late' for him to advise you. He is at the end of a causal 'chain' which begins with your future decision.

At this point it may be objected that if you yourself knew whether the envelope was empty or not, it would surely make sense to take both the envelope and the title-deed and surely knowledge cannot affect the contents of an envelope. To answer this objection, note that a pair of events, one the cause (your actual choice) and the other the effect (God's foreknowledge of your actual choice), is sometimes contingent for its occurrence on a certain condition being met. For example, the pair of events consisting of my scratching a match and the match lighting does not occur if there is no oxygen in the room. In the same way, the pair of events consisting of my planning a surprise for you and my succeeding in carrying out this plan is contingent for its occurrence on your ignorance of the future. Again, Professor Smith growls at Marilyn if and only if he is angry at his wife, only because he does not realize he is being unfair to Marilyn. Thus there is no harm done to the possibility of backwards causation if it is admitted that in certain cases the pair of events, one the cause and the other its previous effect, does not occur if someone has too much information. It may well be that God's foreknowledge of your choice is contingent upon the fact that he has not told you about it. The envelope does have to be sealed.

II

The second argument against the possibility of backwards causation says that causes are supposed to explain their effects. However, if there is backwards causation then there is circular causation and hence circular explanation. But a circular explanation does not explain anything. Thus there cannot be any backwards causation. For example, consider the following case reported by C. G. Jung ('Synchronicity: An Acausal Connecting Principle,' in *The Collected Works of C. G. Jung,* Vol. 8, trans. R. F. C. Hull, ed. Sir Herbert Read *et al.*, London, Routledge and Kegan Paul, 1969, p. 522). A student

> went to Spain, and there, in one of the streets, he recognized the city of his dream. He found the square and the cathedral, which exactly corresponded to the dream-image. He wanted to go straight to the cathedral, but then remembered that in the dream he had turned right, at the corner, into another street. He was curious to find out whether his dream would be corroborated further. Hardly had he turned the corner when he saw in reality the carriage with the two cream-coloured horses

—just as they had been in the dream. Let us suppose that this is a true story and that the future event was actually the cause of the previous dream. Then (1) his turning right at the corner in reality was caused by his remembering that he did so in the dream (together with his curiosity); (2) his remembering that he did so in the dream was caused by his doing so in the dream; and (3) his doing so in the dream was caused by his doing so in reality (assuming, as we are, that the dream was brought about by backwards causation). The possibility of backwards causation thus gives rise to the possibility of circular causation or circular explanation. But this is absurd.

In answer to this objection, first note that when we ask for an explanation of some event E, we are usually satisfied if we are told about another event C which is its cause. We do not usually need to go on and ask for an explanation of C and, if we did, we would not necessarily learn anything more about E. For example, if I ask you why you think Marilyn cares for you and you tell me that Marilyn sent flowers, I do not need to go on and ask why Marilyn sent flowers. If I do, then I may learn only that it was your birthday—which does not tell me more about why you think Marilyn cares for you. Occasionally, however, one is not satisfied merely to know which event gives rise to the event in question, but one needs to know also what gives rise to the cause event. A good student is not satisfied with Kepler's Laws until he has studied Newton and he is not satisfied with Newton's Laws until he has studied Einstein. If one continues to inquire in this fashion, asking about each cause in turn, one runs into either an infinite chain of causes or else a circle. At this point one can ask why there should be such a sequence or circle in the first place. To answer that question one must go outside that sequence or circle. Ultimately, one might introduce a God. God is the ultimate explanans who is himself beyond explanation. For many purposes, however, it is sufficient just to give one cause without giving the cause of the cause. For such purposes there is nothing vicious about circular causality. For example, I can explain why Marilyn came to the party by saying that William came to the party—and this explanation is acceptable even if it is also true that William came to the party because Marilyn came to the party. Or I can explain why Smith is putting a lot of energy into his classes by saying that he has some very enthusiastic students—and this explanation is acceptable even if it is also true that the enthusiasm of the students can be explained by Smith's putting a lot of energy into his classes. It is like the case of a cardboard box whose four flaps are imbricated in such a way that each prevents the one to its right from opening. Flap 1 stays in place because flap 2 stays in place, flap 2 stays in place because flap 3 stays in place, flap 3 stays in place because flap 4 stays in place, and flap 4 stays in place because flap 1 stays in place. All four stay in place because of the way in which they are attached to the box together with the geometry of their imbrication. In the case of Jung's dream, each link in the event-dream-memory-event circle is explained by the previous one. The whole circle must, of course, be explained by something else. Perhaps it was brought about by God.

I conclude that neither Mackie's interference argument nor the circular explanation argument undermines the logical possibility of temporally backwards causation.

Keble College, Oxford

© W. S. Anglin 1981

PEACOCKE ON INTENTIONAL ACTION

By JOHN BISHOP

I

WHAT is the 'right sort of way' in which M's intention to ϕ must cause his doing something he takes to be a ϕ-ing if he is to ϕ intentionally? Christopher Peacocke has recently claimed that there must be a strongly differential explanatory relation between M's intending to ϕ and his matching behaviour: that is, that M's behaviour counts as intentionally ϕ-ing only if (1) M's intending to ϕ is a non-redundant part of the explanation for his behaviour, and this explanation essentially invokes causal laws which specify M's bodily movements as a function of M's intending to ϕ, and (2) the explanatory chain from M's intention to M's movements is such that, at each stage, the explanandum of the previous stage can be recovered from information about the stage itself and its other initial conditions.[1] Intuition suggests that what makes for a deviant causal chain from intention to behaviour is its 'insensitivity' to the content of the intention. Peacocke's 'strong DE condition' appeals as an attempted articulation of what the required sensitivity of behaviour to intention must involve. Yet intuition also raises an immediate objection: in what sense can behaviour be understood as 'the function of an intention'? Peacocke relies on a version of physicalism to surmount this obstacle:[2] the functional relationship between an intention and bodily movements applies to them only under their physical, and not under their psychological, descriptions.

II

Does the strong DE condition supply the missing element in a causal definition of intentional action? To do so, the condition must be sufficient as well as necessary for non-deviancy, and it will be my second aim in this paper to examine whether it is. My first aim, however, is to consider whether the inclusion of the DE condition in a definition of intentional action can meet what I take to be an important constraint on the acceptability of a definition of any concept whose use is already fully mastered. Is there any impediment to supposing that, in applying the concept, naive masters of it are, albeit unawares, employing the criteria articulated in the definition?

---

[1] *Holistic Explanation*, Oxford, 1979. The requirement of differential explanation—part (1) —is based on the 'rough definition' Peacocke offers on p. 66; the requirement of stepwise recoverability—part (2)—is given at p. 80. An event is said 'specifically to reflect' a cause which differentially explains it; causal chains in which each event is differentially explained by its predecessor are said to be 'sensitive' (p. 69).

[2] See *Holistic Explanation*, pp. 69–70. In Part III of this work, Peacocke defends a 'realization' version of the identity theory.

Some causal accounts of the concept of intentional action fail spectacularly to satisfy this constraint. Alvin Goldman, for example (*A Theory of Human Action*, Englewood Cliffs, 1970, p. 62), characterizes the difference between deviant and non-deviant chains as an as yet undiscovered empirical difference between laws which subsume them. Yet, even if such a difference had been discovered, mastery of the concept of intentional action would not require a knowledge of it. And so it is absurd to suggest that, when we judge a given causal chain deviant, what we are 'really' doing is to recognize that the causal law subsuming it is not of the right kind for intentional action. The point is not that we do not, as a matter of fact, understand ourselves to be employing such a criterion—for that would be no problem, since conceptual analysis may bring to light previously unknown characterizations of distinctions with which we have long operated confidently and intuitively. The point is rather that we *could not* understand ourselves as operating with Goldman's criterion, for that would require our agreement that we can apply a purely empirical distinction without knowing what it is. Of course, there probably are significant empirical differences between deviant and non-deviant causal chains. But it is a serious confusion to include in the definition of the concept of intentional action a requirement that the causal link from intention to behaviour should have a certain *but unknown* empirical character.

Peacocke's proposal appears free of this confusion. For him, the concept of intentional action is the concept of behaviour taken by the agent to match the intention which causes it *and* which strongly differentially explains its occurrence. If so, then what we are 'really' doing in sorting deviant from non-deviant instances is to apply a distinction between strongly differential and other kinds of causal explanation. The fact that we do not think of it this way is no objection. Peacocke's account apparently does not imply that masters of the concept of intentional action should possess specialized empirical knowledge which they manifestly lack. One might, plausibly, be able to recognize an explanatory relation as differential or otherwise without having any special empirical knowledge about the precise causal laws involved, just as, for example, we can all be sure that the windscreen shattered because it was hit by a flying stone even though we do not possess the causal explanation for this event in full empirical detail. Peacocke himself draws attention (p. 96) to the fact that the restrictions imposed by the strong DE condition on a causal chain do not employ concepts drawn from the special sciences, and that, accordingly, the results of scientific investigation are not relevant to their specification.

It appears, however, that while further empirical investigation is not relevant to the *specification* of the criteria which, according to Peacocke, underlie our judgments of intentional action, it is relevant to the *application* of these criteria. For, if Peacocke is right, our classification of behaviour as intentional action turns on whether or not the intention concerned strongly differentially explains the behaviour it causes. By his own admission, this question makes sense only with respect to certain physical descriptions of behaviour which no master of the concept of intentional action need know. Consequently, anyone who wished (and for practical purposes this would be a silly wish!) to make conscious use of Peacocke's criteria in judging of intentional action would have to await empirical information about the physical identity of certain mental events. There is thus an obstacle to our understanding ourselves as operating with Peacocke's criterion, just as there was with Goldman's—though it is not quite the same kind of obstacle. The difficulty can be put thus: if Peacocke regards it as a merit of his account that the deviant/non-deviant conceptual distinction is not analysed as an as yet unknown empirical difference, why is he apparently unconcerned that his proposed analysis makes the distinction applicable only in the light of as yet unobtained empirical information? Do not the principles which prompt rejection of an account like Goldman's threaten Peacocke also?

III

I turn to my second aim: is the strong DE condition sufficient for non-deviancy? It seems not. Consider this argument. Peacocke allows (in footnote 12, p. 70) that a state of nervousness might specifically reflect the intention which caused it. The initial implausibility of this is set aside by the descent to the physical. It *could* happen that an intention and a causally consequent state of nervousness were related *under their physical descriptions* in a functional relationship of the kind required for differential explanation. Now, if this is possible, it must surely equally be possible that a nervous state and behaviour caused by it should also be so related. The device of descent to the physical may with equal readiness be employed here. Accordingly, it must be logically possible for the very same feature of an agent's nervousness which specifically reflects his intention *also* to differentially explain his behaviour, where that behaviour happens to match his intention. Apparently, then, there can be causal chains from intention to behaviour which pass through nervousness but which *also* satisfy the DE condition. And this suggests that the DE condition may be satisfied by some deviant causal chains. A specific counterexample cannot, of course, fairly be demanded, since we lack the empirical knowledge needed to illustrate the descent to the physical required by the DE condition.

It may be urged that this argument ignores the condition of stepwise recoverability which will surely not be met by the kind of causal chain envisaged. Surely, even if a nervous state does specifically reflect the intention which caused it, and does differentially explain resultant matching behaviour, this will only ever be in virtue of functions which are many-one, and which therefore preclude stepwise recoverability. It may be replied, however, that there is no reason to exclude the logical possibility that the functions involved should all be one-one, and in that case the strong DE condition would be met.

In short, my complaint is this: our intuitions about intentional action provide no guarantee that the strong DE condition suffices for the non-deviancy of a causal chain from intention to behaviour. Precisely because this condition draws us away from the level of description at which those intuitions operate to the level of realizing physical events which may or may not bear suitable functional relationships to each other, it becomes possible to imagine that a causal chain from intention, through nervousness, to matching behaviour might satisfy even the strong DE condition. Granted, at an intentional-psychological level of description this suggestion hardly makes the least sense: but at the physical level of description it is perfectly intelligible, and threatens a source of counterexamples to the hypothesis of the sufficiency of strong DE for non-deviance.

Peacocke has the following reply available. If there were 'nervousness' of the kind envisaged, then it would count as a genuine if highly unusual part of the realization of an ordinary, non-deviant, intentional action. If nervousness can be realized by a state which is so sensitive a transmitter of information, then it has arguably lost a feature essential for it to count as a state of nervousness. This reply has considerable force. I shall accept it for the present. We may, indeed, be obliged by the general adequacy of the strong DE criterion to revise our intuitions about cases as recondite as those just envisaged. To take the debate further, we may consider how Peacocke responds to a different kind of counterexample to the sufficiency of the strong DE condition.

This counterexample, which Peacocke himself adduces (p. 87), is a familiar case. Suppose that a neurophysiologist can recognize an agent's intention to ϕ from a knowledge of his brain states, and interferes with the agent's brain to produce precisely those neural conditions sufficient to cause him to behave as intended. The strong DE condition could be satisfied in such a case, and yet the neurophysiologist's victim would not be ϕ-ing intentionally.

Some caution is required in treating this example, since, if a certain addition is made to its description, it becomes plausible that the victim *does* ϕ intentionally. If the agent knows of his plight, and truly believes the neurophysiologist to be reliable, he may intentionally bring it about that he moves his arm (say), but only by doing something else, namely, forming the intention to move his arm. To preserve the counterexample we might, therefore, specify that the agent does not have this knowledge.

Alternatively, we may note that, anyway, the victim cannot move his arm (or whatever) as a *basic*[3] intentional action. And we are here concerned with sufficient conditions for intentional action generally, and so, crucially, for basic intentional action.[4] Accordingly, the counterexample stands as it is, having no need of elaboration as to the agent's awareness or otherwise of his situation.

Peacocke's response is disappointing. He concedes that 'the requirement of DE must be filled out to one that also requires that the chain from intention to bodily movement should not run through the intentions of another person' (p. 88). (It is clear from footnote 23, p. 88, that this claim is intended to apply only to basic intentional action.) What is disappointing here? Just that no amount of talk of 'filling out' the DE requirement will obscure the fact that a quite disparate necessary condition needs to be 'tacked on' to the pleasingly unified account Peacocke has so far maintained. The aim, surely, is to supply a theoretical account of the deviant/non-deviant distinction which will make precise, and simultaneously unify, the vague and metaphorically expressed intuitions we have about it. Peacocke seemed to promise to fulfil that aim. Yet he now has to accept an ungainly exclusion condition expressed in the language of our immediate intuitions. The account becomes lopsided. But the damage is not just aesthetic! Peacocke's strategy is under threat. For, in the light of this concession, why should we retain the confidence with which Peacocke began (see, e.g., p. 59) that it is our intuitions about the 'sensitivity' of a causal chain which hold the secret of a clear understanding of intentional action, rather than our intuitions about 'origination', or 'agent-causation'? Perhaps, indeed, these two kinds of intuition stem from something more basic which has not yet been brought to light.

The strong DE condition, then, by Peacocke's own admission, is not sufficient for non-deviancy. However, Peacocke does believe that his condition is sufficient once augmented by a clause excluding chains which pass through the intentions of another agent. I shall argue that this belief is mistaken.

IV

It will be instructive to consider how Peacocke deals with an example due to Pears, who discusses

> a gunman who intends to fire his gun and whose cortex sends out the initiating event: unknown to the gunman, the nerve to his index finger

---

[3] A basic action is an action which an agent does without doing (having to do) anything else in order to do it. The present context does not require any more precise specification of this concept, which is due to Danto ('Basic Actions', *American Philosophical Quarterly* 2, 1965, pp. 141–48).

[4] It is, I believe, relatively unproblematic to provide an account of non-deviant *non-basic* intentional action. See, for example, the suggestions of Goldman (op. cit., p. 59) and Armstrong (D. M. Armstrong, 'Acting and Trying', *Philosophical Papers* 2, 1973, pp. 1–15).

> has been severed, but the intended movement still takes place because the impulse produced by the initiating event attracts some lightning which itself generates the required impulse in the severed section of the nerve (p. 93).

Presumably it is likely that the DE condition will fail in this case, yet Peacocke admits that it *could* be satisfied. In which case, he suggests, 'the lightning is no different from a rather unreliable external aid, of a kind a patient might think at least better than nothing' (pp. 93–94). Peacocke thus makes here the kind of reply I suggested he would make to my case of a causal chain which satisfies the strong DE condition but runs through nervousness: namely, to treat the case as one of genuine intentional action. Is this reply satisfactory?

In order to answer this, let me note that, evidently, every deviant intention-behaviour chain has something unusual about it. But not all abnormalities in such chains introduce deviance: consider, for example, intentional action achieved with the aid of a prosthetic device. It is a matter of interest *at what point* in an intention-behaviour chain an abnormality occurs. In some cases, as the nervousness example illustrates, it is the formation of the intention to ϕ which *itself* initiates an unusual causal chain resulting in accidentally matching behaviour. Such cases are to be distinguished from those in which an intention to ϕ is formed, and the agent acts on this intention (that is, he tries to carry it out—though trying is not, of course, a matter of doing something else first), but the way in which the action is physically realized is unusual. The gunman case may be of this second type, at least if the strong DE condition is satisfied.

The difference between the two types of case can be put thus: in the first, the agent's opportunity to try to carry out his intention is pre-empted. And so, although the agent does exhibit ϕ-ing behaviour, he does not intentionally act on his intention to ϕ even though this intention causes his behaviour. In the second sort of case, however, there is no pre-emption of the agent's opportunity to act. The agent takes that opportunity, and intentionally performs the act, but in an unusual way, and only by the grace of chance.

Clearly, whenever an agent's opportunity to act on his intention to perform a basic action is pre-empted, no intentional *basic* action *of his* occurs, even if what he intended is caused by his forming that intention. Since it is logically possible for the strong DE condition to be satisfied, and yet for pre-emption to occur, the strong DE condition does not suffice for non-deviancy. Indeed, since the strong DE condition can be satisfied in the presence of pre-emption which is *not* due to the causal chain's passing through the intentions of another, the augmented strong DE condition is not sufficient either. (My nervousness case can now be seen as one in which 'internal' pre-emption of the agent's opportunity to act occurs despite the satisfaction of the strong DE condition.) Passing through another's intentions is a typical, but not the only possible, source of pre-emption of the agent's chance to act.

What of cases in which there are causal abnormalities but no pre-emption? Some certainly count as non-deviant. But do they all? Perhaps they do only if the strong DE condition is satisfied. Assuming that there is no pre-emption, the question still remains what forms of taking the non-pre-empted opportunity to fulfil a basic intention count as actually performing the corresponding basic intentional action. Peacocke's strong DE condition may acceptably capture the content of our intuition that what is required is that the agent should have control over the movements which realise the fulfilment of his intention.

V

I have offered three objections to Peacocke's account of the deviant/non-deviant distinction as it applies to intentional action. First, it does not supply sufficient conditions for non-deviancy, even when augmented by the exclusion of chains which run through another's intentions. Second, the admission of this augmentation in the first place casts doubt on Peacocke's crucial strategic assumption—that deviance in intention-behaviour chains has to do with 'lack of sensitivity'. I have suggested that the concept of non-pre-emption of the agent's opportunity to act is also crucial, and cannot be reduced to considerations about sensitivity. Third, Peacocke relies on the descent to the physical to make the concept of differential explanation intelligible. But this is at the cost of rendering it mysterious how naive mastery of the concept of intentional action is possible.

If these objections are well made, the moral is that the deviant/non-deviant distinction is not a matter of a difference in kind amongst causal explanations of behaviour in terms of intentions. And this counts as a serious defect in what may nevertheless still remain our best available account of the concept of intentional action—a causal account.

The University of Wollongong © JOHN BISHOP 1981

DEFINING INTRINSIC VALUE

By RODERICK M. CHISHOLM

IN their paper, 'Isolating Intrinsic Value', (ANALYSIS, January 1981, pp. 51–3) Bodanszky and Conee express doubts as to whether an "isolation approach" will yield adequate definitions of intrinsic value and related concepts. I think that they are right in their criticism of the particular attempt that they discuss. But I am quite certain that they are wrong in concluding that no such "isolation approach" yields adequate explications of these concepts.

I believe they assume, as I do, that the bearers of intrinsic value are states of affairs—those abstract objects, designated by that-clauses and sentential gerundives, which are sometimes called "propositions". *Someone experiencing pleasure,* for example, would be a state of affairs that is intrinsically good and *someone feeling displeasure* a state of affairs that is intrinsically bad.

Perhaps they also agree with me that the intrinsic value of a state of affairs, as contrasted with its instrumental value, is a feature that that state of affairs has necessarily. If a state of affairs is intrinsically good, then it is intrinsically good in every possible world in which it obtains (or is true). But a state of affairs that is instrumentally good need not be instrumentally good in every possible world in which it obtains (or is true). And analogously for badness and indifference.

It can be shown, I think, that all intrinsic value concepts may be analysed in terms of intrinsic preferability. In the definitions that follow 'intrinsically' should be thought of as implicit throughout.

A state of affairs p is *the same in value* as a state of affairs q, if and only if p is not preferable to q and q is not preferable to p. A state of affairs is *indifferent* if and only if it is the same in value as its negation. A state of affairs is *good* if and only if it is preferable to a state of affairs that is indifferent. And a state of affairs is *bad* if and only if a state of affairs that is indifferent is preferable to it (compare Roderick M. Chisholm and Ernest Sosa, 'On the Logic of "Intrinsically Better",' *American Philosophical Quarterly,* Vol. III, 1966, pp. 244–49).

The question, then, turns upon the analysis of the locution 'p is intrinsically preferable to q'. What can we say about it?

Following Brentano and Meinong, I would suggest that intrinsic value concepts may be defined in terms of the appropriateness of certain intentional attitudes. For certain attitudes may be said to be *appropriate to*—and indeed *required by*—their objects (compare Franz Brentano, *The Origin of Our Knowledge of Right and Wrong,* London: Routledge & Kegan Paul, 1969, and Alexius Meinong, 'Zur Grundlegung der allgemeinen Werttheorie' and 'Ethische Bausteine', in *Abhandlungen zur Werttheorie,* in *Meinong Gesamt Ausgabe*, Vol. III, Graz, Akademische Druck- u. Verlagsanstalt, 1968). Thus a pro-attitude is appropriate to a state of affairs that is intrinsically good, and an anti-attitude is appropriate to a state of affairs that is intrinsically bad. How, then, are we to define intrinsic preferability?

If a state of affairs p is intrinsically better than a state of affairs q, then it is appropriate to *prefer* p to q. And the contemplation of the two *requires* that p be preferred to q. Hence we could explicate intrinsic preferability this way:

> p is intrinsically preferable to q $=_{df}$ p and q are necessarily such that, for any x, the contemplation of just p and q by x requires that x prefer p to q.

One might qualify the definition by saying: '. . . the contemplation of just p and q *as such* by x requires that x prefer p *as such* to q.' One would thus be using the 'as such' to stress the fact that the contemplation of a state of affairs *as such* is the contemplation just of *that* state of affairs—as distinguished, for example, from the contemplation of some wider state of affairs which one may think that the given state of affairs brings along with it. And one may say 'x prefers p as such to q' in order to stress the fact that x is ranking just *those* states of affairs, as distinguished from any wider states of affairs that may include them. But I will assume that the reader takes the definition literally: 'the contemplation of just p and q' would not refer to the contemplation of any wider state of affairs that might include p or q; and 'x prefers p to q' would not refer to a ranking of any states of affairs that are wider than p or q.

It may be noted that, in ranking p and q as such and thus 'in isolation,' one is concerned just with those features of p and q that they have necessarily and not with any particular features that they may happen to have in this world but not in others. And so one might say that the states of affairs are objects of the "antecedent will" as distinguished from the "consequent will".

This approach to intrinsic value, then, makes use of the concept of *requirement*. And the concept of requirement, there is reason to think, is the central concept of ethics (compare Roderick M. Chisholm, 'Practical Reason and the Logic of Requirement', in Stephan Körner, ed., *Practical Reason,* Oxford, Basil Blackwell, 1974; and Philip C. Quinn, *Divine Commands and Moral Requirements*, Oxford, The Clarendon Press, 1978). The "isolation approach", therefore, not only yields adequate definitions of the basic intrinsic value concepts, but has the following advantage as well: it provides a way of reducing the concepts of the theory of value ("axiology") to those of ethics ("deontology").

Brown University

© Roderick M. Chisholm 1981

ABSOLUTISM v CONSEQUENTIALISM: NO CONTEST

By Don Locke

CONTEMPORARY moral philosophy has made much of a distinction between Absolutism and Consequentialism, which appears to be the direct descendent of a more traditional distinction between Deontological and Teleological ethics. Since these distinctions are themselves unclear, it is equally unclear whether they are the same distinction. But neither, I shall argue, marks a useful, or even a genuine, contrast. The choice between Absolutism and Consequentialism, often presented to us, is in fact no choice at all: it is possible to have your cake and eat it, or, as I would prefer, to reject them both.

The term 'Consequentialism' seems apt enough, and certainly to be preferred to the increasingly wide and loose sense of 'Utilitarianism' to refer to any ethic which appeals to consequences. The import of 'Absolutism', similarly, is clear and familiar. So on the one hand we have theories which base moral value on the consequences of our actions; on the other we have theories that some actions have a particular moral value whatever the consequences (I concentrate on actions only for simplicity, and not because I believe that only actions can have moral value). But these are not, strictly, opposite hands. The phrase 'whatever the consequences' generates the illusion that they are, but replace it by the (at least) equally correct 'whatever the circumstances', or, more broadly still, 'whatever else may be the case', and the illusion disappears. The paradigm example of a Consequentialist ethics, Classical Utilitarianism, is also a perfect example of Absolutism: any action which maximizes pleasure or happiness is right and good absolutely, whatever else may be the case. And if some ethical positions are simultaneously Consequentialist and Absolutist, others are neither. A theory of prima facie duties and obligations like that of Ross, which holds, for example, that the keeping of promises is good in itself, good because it is a promise and not because of the consequences, but can sometimes be wrong because of other features of the situation, is equally clearly neither Consequentialist nor Absolutist. And no one, I imagine, would accuse Bernard Williams of being either a Consequentialist or an Absolutist.

The obvious contrast with Consequentialism is not Absolutism but 'Intrinsicalism', the theory that (at least some) actions have a moral value because of their intrinsic nature, and not just because of their consequences, whatever they might be. Strictly, perhaps, these are not contradictories, but contraries: a theory which judged actions by the motives with which they were done would be neither Intrinsicalist nor Consequentialist, though no doubt the reference to motives would in turn involve a reference to the supposed nature or consequences of the intended action. The obvious alternative to Absolutism is Relativism, but since that term already has other connotations in ethics, I will prefer the clumsier 'Situationalism', the theory that the moral value of actions depends on, and can vary with, the situation.

The contemporary conflation of these two, clearly different, distinctions can perhaps be traced back to the introduction of the term 'Consequentialism' in G. E. M. Anscombe's much-discussed article, 'Contemporary Moral Philosophy'. Anscombe rejects the Consequentialism of British ethics since Sidgwick, and favours instead a more traditional Hebrew-Christian ethic which involves 'the prohibition of certain things simply in virtue of their description as such and such identifiable kinds of action, regardless of any further consequences' (*Philosophy*, 1958, p. 10). Not unnaturally this alternative has since become known as Absolutism, but as Anscombe describes it, it is in fact a combination of Absolutism and Intrinsicalism, and, I suspect, the latter is more characteristic of traditional European ethics than the former. The two are more carefully separated in Peter Geach's 'The Moral Law and the Law of God' (*God and the Soul*, Chap. 9, espec. pp. 119–20), where he argues, in effect, that whereas all sane and honest people are Intrinsicalists ('I am not saying a sane and honest man must think one should *never* lie; but I say that, even if he thinks lying is sometimes a necessary evil, a sane and honest man must think it an evil'), it requires a belief in God, and what is more a belief of a certain sort, to be an Absolutist, or rather an Absolutist and Intrinsicalist combined ('what I mean is that certain sorts of action are such *bad things to do* that they must never be done to secure any good or avoid any evil').

Nevertheless the failure of Anscombe, and her critics, to separate Absolutism from Intrinsicalism has meant that those who wish to defend Utilitarianism, in particular, have felt obliged to attack Absolutist ethics, while those who side with Anscombe, if only in part, have felt obliged to defend it. So we find Thomas Nagel, for example, distinguishing between Utilitarianism, which 'gives primacy to a concern with what will *happen*', and Absolutism, which 'gives primacy to a concern with what one is *doing*' (*Mortal Questions*, p. 54). But the latter, surely, is Intrinsicalism rather than Absolutism, a confusion which makes Nagel's defence of the view that there are other factors which not merely have to be weighed along with the consequences, but may even outweigh them, more convoluted than it need otherwise have been. And eventually we find the confusion occurring even without the terminology. Elazar Weinryb, for example, contrasts Consequentialism with Formalism, 'according to which the moral character of an action is determined by its own nature'; and then assumes, without further argument, that the crucial difference between the two is captured by the principle that 'some actions are such that it would always be wrong to perform them, whatever the consequences of not doing so' ('Omissions and Responsibility', *Philosophical Quarterly*, 1980, p. 4).

However, this conflation of Absolutism and Intrinsicalism might also be traced back to the earlier terminology of Deontological and Teleological ethics. So far as I can discover this contrast was introduced into philosophy by C. D. Broad, in an attempt to avoid an ambiguity in Sidgwick's use of 'Intuitionism'. Certainly that arch-Teleologist Bentham, writing almost a hundred years earlier, was so innocent of it that he actually titled one work *Deontology*, defining the crucial term as 'the science by which happiness is created out of motives extra-legislatorial' (I, ii, 28). The terms themselves suggest a useful distinction between those things we ought, or ought not, to do because we have specific duties, and those things we ought, or ought not, to aim at, but where no particular duties are involved—a distinction which will, of course, collapse if we happen to believe there is always a duty to aim at the good and avoid the bad. Or, at a meta-ethical level, we might distinguish between those theories which take 'ought' as the fundamental moral notion, and define 'right' and 'good' in terms of it, and those which take 'good' as the fundamental notion, and define 'ought' and 'right' in terms of that. Something of this sort seems to lie behind Frankena's distinction: 'A teleological theory says that the basic or ultimate criterion or standard of what is morally right, wrong, obligatory, etc., is the nonmoral value that is brought into being. . . . Deontological theories . . . assert that there are, at least, other considerations which may make an action or rule right or obligatory . . . other than the *value* it brings into existence . . .' (*Ethics*, pp. 13–14).

But this is not how Broad himself introduced the distinction. Instead he first identifies the crucial question as 'whether some types of action are *intrinsically* right or wrong or whether the rightness or wrongness of actions always depends on their conduciveness to certain *ends*', which seems to be the distinction between Intrinsicalism and Consequentialism. But Deontological theories are then presented as Absolutist in form, holding 'that there are ethical propositions of the form: "Such and such a kind of action would always be right (or wrong) in such and such circumstances, no matter what its consequences might be" (*Five Types of Ethical Theory*, p. 206). Whether Teleological theories are also Absolutist, in Broad's opinion, is not made clear. But we think, once again, of W. D. Ross, the paradigm Deontologist in the eyes of many (though Ross himself seems not to use the term), yet by Broad's definition not a Deontologist at all!

I suspect that the cause of this repeated failure to separate Intrinsicalism from Absolutism is the failure to draw one of many possible distinctions between the right and the good, the wrong and the bad. An action may be bad and yet, in the circumstances, still the right action, inasmuch as it is the lesser of two evils: we have, as we put it, to make the best of a bad job. And similarly, according to some of the more demanding ethics at any rate, an action may be good and yet still not the right action, because an even better one is available. Now if we draw this distinction we can say that if, according to the Intrinsicalist, certain sorts of action are good or bad simply because of the sort of action that they are, then any such action will be good or bad, at least to that extent, regardless of what else may be the case. It does not follow, however, that such actions are always right or wrong, are right or wrong absolutely. That will depend on whatever other factors, besides the intrinsic value of the action as such, have to be taken into account. So we could say, confusingly, that the Intrinsicalist must regard certain types of action as absolutely good or bad, without necessarily regarding them as absolutely right or wrong. I prefer to avoid this source of confusion by restricting Absolutism to the latter position, which seems to be the natural interpretation.

Clearly, then, we have two different distinctions, where previously we had one: between Consequentialism and Intrinsicalism on the one hand, between Absolutism and Situationalism on the other. But these distinctions are themselves not entirely clear. The distinction between Intrinsicalism and Consequentialism depends crucially on a distinction between act and consequence, but as is now well known, there is no absolute difference here. Whether something counts as part of the intrinsic nature of an action, or as a consequence, depends on how the action is described: if we call it killing then, no doubt, it is intrinsically bad; but if we describe it as firing a gun, with the result that someone dies, then the evil lies not in the act but in its consequences. So, unless the Intrinsicalist is prepared to restrict moral value to those basic acts (if there are any) which cannot be represented as a consequence of what we do, or the Consequentialist is prepared to restrict moral value to those consequences (if there are any) which can never be incorporated into a redescription of the action—and neither view seems particularly tempting—it seems that, with suitable juggling, an Intrinsicalist can always be turned into a Consequentialist, and a Consequentialist into an Intrinsicalist. Thus the Utilitarian, for example, does regard one sort of action as intrinsically good, good simply in virtue of its description as such and such an identifiable kind of action, viz. maximizing human happiness. Indeed this is just the sort of action which typically turns up on Deontologists' lists of intrinsically good behaviour! (These points, and many others undermining the Intrinsicalist/Consequentialist distinction, are made in an excellent article by James Griffin, 'Consequences', *Proceedings of the Aristotelian Society*, 1964–65, which deserves to be much better known than it is, especially for its careful disentangling of Consequentialism from what Griffin terms Naturalism.)

There is, however, another possible way of distinguishing, morally, between act and consequence (I owe this suggestion to Christopher Kirwan). An action consists in the bringing about of various states of affairs. On the one hand we might say that the value of an action consists simply and solely in the value of those states of affairs it brings about, or prevents. Such a view might well be called Consequentialism. But on the other hand we might say that moral value attaches also to the fact of bringing such states of affairs about, or preventing them. Someone who says, for example, that violence is such an evil that it is wrong to use violence even when the result would be less violence in the world, is not necessarily contradicting himself (as Narveson suggests, 'Pacificism: A Philosophical Analysis', *Ethics*, 1965). He may believe that although the existence of violence is an evil, it is worse still to do violence, to the extent that no reduction in the total amount of violence can outweigh the fact that you have used violence yourself (cf. T. Regan, 'A Defence of Pacificism', *Canadian Journal of Philosophy*, 1972–73). But such a view is not very aptly called Intrinsicalist. It is, rather, an aspect of the conflict between Utility and Integrity, presented to us by Bernard Williams ('A Critique of Utilitarianism', in J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*).

The other distinction, between Absolutism and Situationalism, also tends to evaporate under scrutiny. If the Situationalist is to escape the charge of sheer inconsistency, he must allow that if an action is right or wrong in one situation, then it will also be right or wrong in any other situation which is either exactly or relevantly similar to the first: if an action is right, or wrong, in one situation but not in another, there must be some morally relevant difference between the two. Hence once we specify all the morally relevant features of the particular situation, we will have specified something which, even for the Situationalist, is absolutely right or wrong, right or wrong whatever else may be the case.

But here too there may be a difference, even if it turns out not to be the difference it first appeared to be. Typically the Absolutist believes that it is possible to specify, in advance, certain things which one must do, and certain things which one must not do, whatever else might be the case. The Situationalist, on the other hand, does not believe that this is possible, in advance of the particular situation. Certainly if an action which was wrong in one situation turns out to be right, or even alright, in another, there must be some difference between the two. But according to the Situationalist it is only by considering the concrete situation that we can discover whether there is some such crucial difference, how the various factors affect the rightness or wrongness of the action in question, or even whether they are morally relevant in the first place. It would be tempting to label these two approaches not Absolutism and Situationalism, but Ethical A Priorism and Ethical A Posteriorism respectively, if those terms did not already have rather different epistemological connotations.

To put it another, possibly different, way, moral evaluation is, for the Situationalist, essentially open-ended: there is no telling whether some further factor may not come to light which will force him to alter or adapt any generalization which has held true so far. But for the Absolutist, moral evaluation can be closed, and closed in advance: there are some factors such that, once they apply, no other considerations can make a relevant difference. It appears that an Absolutist ethics will have either to be monistic, reducing all morality to a single value such as the maximization of happiness, so that once that is assessed, nothing else remains to be considered; or if it is pluralistic, it will have to be rigidly heirarchical, with some values inevitably taking precedence over others so that, once again, once they have been assessed, nothing further needs to be considered. And, I think, vice versa: an ethics which is either monistic or rigidly heirarchical in its values and principles, will inevitably be Absolutist, in this sense. A Situationalist ethic, on the other hand, can afford to be more messy: the relative precedence of different values may vary from case to case: in one situation we might value truth-telling above beneficence, in another we might value beneficence above truth-telling. And although the former approach may commend itself to the neat and tidy minds of those philosophers (and, temperamentally, I am one) who would prefer to present us with a system of morals rather than a hotch-potch of moral intuitions, the latter seems closer to moral thought as actually practised. It is not just that each of us tends to regard a complex variety of considerations, not all of which can be reduced to others, as morally relevant, where these considerations can come into conflict in different ways in different contexts. These considerations may also interact with each other, and with other features of the situation which may not be regarded as of prime moral concern in themselves, in a way which, ultimately, makes it impossible to order our values outside of the particular concrete situations to which they apply.

There are, therefore, distinctions to be drawn, but they are not quite as the terms Intrinsicalism and Consequentialism, or Absolutism and Situationalism—much less the false dichotomy between Absolutism and Consequentialism—might imply. But such is its convenience as a way of indicating the distinction between broadly Utilitarian and non-Utilitarian ethics, the Absolutist/Consequentialist terminology will no doubt survive in spite of me. To that I can say only that any appearance of greater accuracy or precision, in comparison with the terms 'Utilitarian' and 'non-Utilitarian', is entirely an illusion.

University of Warwick © Don Locke 1981

A UTILITARIAN PARADOX

By FREDERICK KROON

I

CONSIDER the following scenario. A person I am observing, Jones, is inflicting pain on someone. Call his action 'A'. I, a confirmed utilitarian, am about to pass moral judgment on action A. Under normal circumstances I would class A as terribly wrong, but on this occasion I know that others are aware of the fact that I am about to pass moral judgment on A, and I also know that these others, who in some ways are superior to me (they can directly intuit, by a kind of mental telepathy, a person's emotional responses, including his moral approval or disapproval of actions), will learn from my moral stance: if I evaluate A as wrong, that will reinforce their hatred of such acts (a *very* good thing, as there are many of these 'witnesses' to my act of evaluation), while if I evaluate A as right, that will reinforce their admiration of such acts (a *very* bad thing). Suppose I know that there are no other morally relevant consequences, and suppose I cannot communicate with these 'witnesses'. Suppose, in fact, that this is the final moral 'training' that I, or anyone else, can provide for these witnesses, who are due to spend the rest of their lives on a desert island, applying the moral lesson they will presently learn from me (with horrendous consequences, should I approve of cruelty like Jones').

What should I do? A negative evaluation—the one that most swiftly comes to mind because of the suffering of Jones' victim—will result in very good consequences (the evaluation's effect on the witnesses to my evaluation), and hence A will include in its total field of consequences beneficial consequences that outweigh the suffering of the victim (and that also, let us suppose, clearly outweigh the good consequences of other actions open to Jones). A negative evaluation is known by me to be incorrect, therefore, and so I cannot judge that A is wrong. (Recall that the witnesses to my evaluation can detect my mental set, and so I cannot pretend to be horrified at A while really approving of A.) Nor can I believe that A is right, since that evaluation is also known by me to be incorrect: a positive evaluation, and hence the action A that evokes the evaluation, leads to very bad consequences, and hence is known by me to be wrong. Should I perhaps rate the action 'morally indifferent', or refuse to commit myself? Scarcely, for that (let us suppose) is known by me to encourage behaviour similar to A on the part of the witnesses to my evaluation, or non-evaluation. Hence in that case A will be known by me to be wrong, and so I can scarcely affect moral indifference, or refuse to commit myself on the question of A's morality.

Unlike the person who observes the occurrence of A, judges A to be *wrong* (ceteris paribus) and subsequently discovers that the effects of that judgment make A in fact *right* (contra his initial moral estimate), I am in trouble. I know too much about the effects my moral evaluation of A will have, and so I cannot indulge in *ceteris paribus* moral judgments, later on to be revised when further information becomes available. As a rational utilitarian, I naturally issue sincere moral judgments that take full account of *all* the expected consequences of the action being judged, and not just of its immediate consequences. My problem, in short, is this. Confronted with A, and the question of how I should judge A, I am unable to proceed, for I have a string of counterfactuals whose structure precludes the making of *any* evaluation: were I to evaluate A as (1) wrong, (2) right, (3) indifferent, A *would* be (1) right, (2) wrong, (3) wrong. I cannot feel genuine moral horror in the face of A since I know that such a judgment on A will suffice to render that judgment entirely inappropriate. And, similarly, I cannot feel genuine moral pleasure in the face of A since I know that such a judgment on A will suffice to render my judgment entirely inappropriate. Nor can I pretend indifference. It is my state of knowledge, in every case, that makes it impossible for me to arrive at an honest utilitarian appraisal of action A[1].

II

This is a utilitarian 'paradox', here couched as a problem for act-utilitarianism but also a problem (I suspect) for some familiar versions of rule-utilitarianism.[2] Paradoxes, however, always presuppose inbuilt assumptions whose articulation, and consequent rejection, may lead to theoretical sanity once more. The present paradox is no exception. In fact, as is often the case, numerous suggestions for drawing its sting exist, not all of them worth taking seriously. One 'way out' challenges the view that beings could exist who can discern moral approval and disapproval without understanding the intellectual route that leads to these sentiments. For once this view—a kind of Humean assumption—is rejected, my witnesses must be considered competent to discover that their rule of action is at bottom paradoxical (like the rule of action in the 'Hangman's paradox'), and hence they must be considered capable of seeing the need for a more rational way of ordering their actions. I reject this 'way out'. I see no confusion in the idea of beings who have access to a person's emotional responses (in particular his moral ones) but not to the full range of his mental resources.

---

[1] Of course I can judge that, *were* A to have consequences that did not extend beyond the immediate ones (suffering to Jones' victim, etc.) then *A would* be wrong. But this is not a moral evaluation of A at all. It is counterfactual implying the appropriateness of a certain evaluation in a certain counterfactual situation.

[2] I owe to Dr. Christine Swanton the observation that it seems *not* to be a paradox for a utilitarianism that adopts a prescriptivist approach—of a rather literal sort—to moral language: for if 'A is wrong' is in reality a directive meaning something like 'don't choose A', then the judgment 'A is wrong' seems the only rational judgment for a utilitarian to make. If this is indeed the case, the paradox arises only on *certain* analyses of moral language.

As a matter of fact, our scenario is not, in its conception, so very improbable. It is doubtful that we really need to appeal to telepathic 'witnesses' to generate our paradox. Moral stances, like other emotional states, manifest themselves in external behaviour: we speak of a person's evincing moral 'shock' and 'disgust', and that does *not* describe a purely internal state, nor a complex state of a person whose behavioural component bears a purely contingent relation to its internal component. So our paradox can be reformulated in terms of my ordinary barely feigned moral disgust/pleasure at an action like A. We can suppose that my witnesses are expert at 'reading' moral disgust/pleasure (most of us are good at this, at least to a degree), that I am a utilitarian who, like most of us, cannot altogether feign his moral emotions, and that my witnesses, on reading my moral reaction to A, have their own inclination to perform like actions reinforced (positively or negatively) in the way described in Section I. (Young children often have their behaviour reinforced along such lines, surely.)

But another defence of our scenario is possible. Acknowledge that it is at least possible that there are such witnesses to a moral evaluation, no matter how improbable the supposition of the effects of my evaluation on their actions, and we can reconstruct the scenario as follows. I am a utilitarian moralist who is unwaveringly convinced, no doubt wrongly, that there are such witnesses to my evaluation of A, and who, in the light of this belief, cannot coherently evaluate, or fail to evaluate, A. This reconstructed scenario no longer depends on the actual existence of such strange witnesses, but its effect is the same. Note that by agreeing that it is at least logically possible that there are such witnesses we are prevented from rejoining that it is the incoherence of the supposition that there are such witnesses that produces the 'paradox'.

A more promising suggestion is this. While the good consequences of my evaluation, if the evaluation is one of disapproval, make the evaluation morally commendable, they do not make the action which is being evaluated morally commendable. An act of evaluation cannot properly be said to be a causal *consequence* of the action being evaluated, although it is, in its own right, an action that has consequences and can be designated as good or bad on the basis of these consequences. Once this is seen, the paradox disappears. Instead, we have a scenario in which the correct, or appropriate, moral reaction (disapproval) to an action A is one that has best consequences.

This response is incomplete. We still need to know why a person's moral reaction to an event cannot be classified as one of the *consequences* of that event. (Presumably the reaction occurred *because* of the event, would not have occurred had the event not occurred, etc.) There are good reasons, in fact, for affirming the classification. We quite naturally say such things as: that action *horrified* us; that is, the action brought about a state of horror within us. Causal terminology seems entirely appropriate when describing how actions affect us morally, even though moral reactions are often only the end-product of a process of reasoning —a calculated weighing-up of perceived and expected consequences in the case of the utilitarian moralist. And if, in spite of this argument, we continue to think it odd to regard one's moral reaction to an action as a *consequence* of the action, perhaps because, in large part, it is a deliberative act that need not be played against the background of an unalterable set of moral principles, it should be noted that utilitarians themselves typically treat the third-person case differently. An action that 'causes' riots is regarded by utilitarians as wrong, *ceteris paribus*, in large part because its causal field includes all the bad effects of riots: injuries, social instability, etc. The kind of behaviour that constitutes a riot, however, is rather more obviously the product of deliberation than are bare moral reactions of the type featured in our own scenario. So, on pain of inconsistency, our utilitarian cannot say that it is the presence of an element of deliberation that disqualifies moral reactions from inclusion in the causal field of the actions they are reactions to.

Nor are we able to take refuge in the view, accepted by some utilitarians, that X is a consequence of action Y only if the occurrence of X is *intended* by the agent of action Y. By adding the condition that Jones performed A because, knowing of the presence of a confirmed utilitarian and telepathic witnesses to that utilitarian's moral reactions, he wanted to determine how the utilitarian would judge A in the light of the presence of these witnesses, and how the witnesses in turn would react to A in the light of that judgment, we are able to reconstruct the scenario in a way that preserves the air of paradox even while conceding, for the sake of argument, the *intentional* view of consequence.[3]

III

Let us reflect for a moment on the moves that generate the 'paradox'. It seems that my attempt at a moral evaluation of A can never get off the ground. And the reason for the breakdown is that we are allowing my evaluation to be based on consequences of the evaluation itself. In this the example is like familiar semantic paradoxes: a sentence like (i): '(i) is not true', if assessed as *true*, has then to be assessed as not true, while, if assessed as *not true*, has then to be assessed as true. Here we are allowing the semantic assessment to be determined by the content of the assessment itself, which in this case (but not always; cf. (ii): '(ii) is true') leads to paradox. It is the logic of our ordinary standards of semantic assessment, and of the utilitarian's standard of moral evaluation, that generates the possibility of such 'ungrounded' acts of assessment and evaluation, and the consequent danger of paradox.

---

[3] Although this response to our paradox may eventually succeed if it insists that a larger number of features of an event X be intended by the agent of Y if X is to be a *consequence* of Y.

A solution that has recently been proposed to draw the sting of semantic paradoxes seems also appropriate in the present case. Tyler Burge[4] argues for the idea that 'true' contextually shifts its extension, where the contexts in question are contexts of semantic *evaluation*. When we evaluate or assess (i): '(i) is not true' as not true (because of paradox), we are evaluating it as not true_i (for some index of evaluation i), understanding the 'true' in (i) as having the same index i. At a second index of evaluation j, however, we can scrutinize (i) anew and agree that (i) is nonetheless true—true_j, that is—since it says of itself that it is not true_i. The details of Burge's proposal need not involve us here. What is important is that the idea of contexts of evaluation can easily and apparently quite naturally be extended to the moral domain. Hence the following proposal. A person's utilitarian evaluation of an action or state of affairs occurs at a context of evaluation. Evaluation conducted at a context of evaluation results in a person's fallible theory of the morality of the action or state of affairs as determined at that context: fallible because a person may not know all the relevant consequences of the action she is evaluating. The concept of 'relevant consequence' is restricted, on the present proposal, to include only consequences that do not depend on the action of evaluation itself; in that sense, we insist on properly *grounded* evaluation. Nonetheless, the action that is evaluated as right or wrong (or indifferent) at *one* context of evaluation, i, may be evaluated anew by the same person or by others. At the *new* context of evaluation, j, the very same action may receive an evaluation whose content (right, wrong, indifferent) may be different, even when there has been *no* reassessment of the perceived and expected consequences already taken into account at context i. There may be a *new* evaluation precisely because at a new context of evaluation consequences of the action that depend on the original evaluation can now be taken into account, now count as *relevant* consequences of the action. Thus my utilitarian reaction of disapproval of action A at context of evaluation o may be entirely appropriate: A is wrong_0, from the point of view of my present knowledge of expected consequences (which, in turn, are consequences from the perspective of context o; call these 'consequences_0'). Once I evaluate the action anew, however, now taking into account the fact that my action of evaluation at context o is itself a consequence of A, I may decide that A is right after all; that is, I may judge that A is right_1, 'right' as evaluated at a context where we are entitled to take account not just of consequences$_0$ but also of the fact that the action of evaluating A as wrong$_0$ itself is a consequence—a consequence$_1$—of A which generates a measure of utility/disutility for A.

---

[4] T. Burge, 'Semantical Paradox', *J. Phil.* LXXVI, 4 (April 1979).

Unfortunately, however, the formal similarity of this little proposal to Burge's proposal for semantics ends at the point of introduction of the idea that 'right/wrong' shift their extension contextually. Whereas for Burge, a sentence that is true$_i$ is also true$_j$ for $j>i$, on the present proposal an action B may be wrong$_0$, right$_1$, wrong$_2$, right$_3$, and so on. (It is a simple matter to construct cases with this property, always supposing, of course, that acts of utilitarian evaluation can, in these cases, be classed as *consequences* of the evaluated action B.) Hence we no longer are guaranteed eventual stability in our evaluations. Worse, it follows that a *usable* concept of right/wrong (the kind of concept *we* use when evaluating an action as right/wrong) is a contextually shifting concept, and hence we, who use such a concept, can arrive at no non-contextual, absolute utilitarian evaluations of actions, even though there is no formal contradiction in the view that an action is nonetheless absolutely right or wrong since, *sub specie aeternitatis*, it may well be a matter of plain descriptive fact that the overall utility of a particular action, measured by its total outcome (including the contribution of acts of evaluation induced by the action), is higher than that of its alternatives.[5]

Alternatives to the view that regards 'right' and 'wrong' as contextually shifting in extension for the utilitarian have their own pitfalls. Thus we have already emphasized that the view that acts of evaluation are *not* consequences of the evaluated action in situations like that of our scenario appears at odds with standard ways of understanding the concept of utilitarian consequence. We are forced to conclude, I believe, that the 'paradox' poses a genuine problem for utilitarianism. While we have canvassed only a few proposals for drawing its sting, difficulties in these proposals for standard utilitarian ideas suggest that some of these ideas, at least, need to be revised or discarded if utilitarianism is to stand as a minimally coherent moral theory.

University of Auckland © FREDERICK KROON 1981

---

[5] Such a divergence between utilitarianism viewed from the outside and utilitarianism viewed from the inside has, of course, also been argued for on other grounds: cf. Bernard Williams' view (e.g., in *Morality*, CUP, 1976: p. 112) that utilitarianism ought, on its own terms, not to be believed by people at large since such belief does not produce best consequences, and the views of those who hold that act-utilitarianism is self-defeating, or, at least, incapable of justifying group-action (even when group-action *does* produce best consequences).

3 JUL 1981

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £8.00 (inland), £9.80 (overseas), $24.00 (U.S.A.), $28.00 (Canada); for individuals £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A.), $17.50 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. **Articles submitted for publication should be addressed to Mr. Christopher Kirwan, Exeter College, Oxford OX1 3DP.**

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

ISSN 0003-2638

PRINTED IN GREAT BRITAIN BY BURGESS & SON (ABINGDON) LTD., ABINGDON, OXFORDSHIRE

ol. 41 No. 3 (New Series No. 191) June 1981

ANALYSIS

Edited by
CHRISTOPHER KIRWAN

CONTENTS

Five o'clock on the sun — J. L. MACKIE
Is this a proof I see before me? — NEIL TENNANT
Reply to Tennant — COLIN McGINN
Is epistemic preferability transitive? — ROY A. SORENSEN
On an argument of Peacocke's about physicalism and counterfactuals — N. M. L. NATHAN
'All actions occur inside the body' — E. J. LOWE
The transitivity of causation — GERMUND HESSLOW
The impersonal formulation of the Cogito — KIM DAVIES
Descartes, Welbourne, and indubitable beliefs — ROBERT FAHRNKOPF
Memories and the world — D. H. M. BROOKS
Escaping the speech act dilemma — KENT BACH
Divine determinateness and the free will defence — DAVID L. PAULSEN
Carr on beliefs and dispositions — PETER SMITH
Killing people intentionally, by chance — KIM DAVIES
Killing, letting die, and euthanasia: a reply to Holly Smith Goldman — PHILIPPA FOOT

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003–2638

Philosophical Studies

An International Journal for Philosophy in the Analytic Tradition

Editor-in-chief: JOHN POLLOCK

Associate Editor: KEITH LEHRER

Philosophical Studies, a journal devoted to work in analytic philosophy, publishes articles in all general areas of philosophical investigation, philosophy of science, epistemology, metaphysics, ethics, and so forth, including the history of philosophy. The main criterion of articles published is that they contain rigorous argumentation and precise analysis of key concepts. Articles in the journal are intelligible to philosophers whose expertise lies outside the subject matter of the article. It is intended that a diligent reader of the journal will be kept informed of the major problems and contributions of contemporary analytic philosophy.

Subscription Information: 1981, vols. 39–40 (8 issues). Private: Dfl. 42,– / US $ 21.00 per volume. Institutional: Dfl. 131,– / US $ 69.00 per volume.

Private subscriptions must be sent to the publishers.

D. Reidel Publishing Company

P.O. Box 17, 3300 AA Dordrecht, The Netherlands
190 Old Derby St., Hingham, MA 02043, U.S.A.

The Philosophical Quarterly

Special Numbers:
The Editors intend to devote one issue in 1982 to articles on **Scientific Realism** and one issue in 1983 to articles on **Medical Ethics.** Contributions should be submitted by **31 December of the preceding year,** to the Executive Editor, The Philosophical Quarterly, The University, St Andrews, Scotland.

Contents of Volume 31, No. 124, July 1981

Knowledge, Belief and Understanding — *R. G. Franklin*
The *Grundlagen* in Frege's Development — *David Bell*
Can Rights be Abused? — *Frederick Schauer*
Fatalism and Causal Determinism — *Michael J. White*

DISCUSSIONS
  Events and Properties — *P. M. S. Hacker*
  Gewirth's Generation of Rights — *Loren E. Lomasky*
  Reference and Speaker's Meaning — *Rod Bertolet*

CRITICAL STUDY
  D. Wiggins, *Sameness and Substance* — *Harold Noonan*

BOOK REVIEWS — BOOKS RECEIVED

Annual Subscriptions:
Individuals £6.00 ($15.00) Institutions £12.00 ($30.00)
for orders placed directly with the Journals Manager, Scottish Academic Press Ltd., 33 Montgomery Street, EDINBURGH EH7 5JX, Scotland.

ANALYSIS 41: 3 JUNE 1981

FIVE O'CLOCK ON THE SUN

By J. L. MACKIE

> "BUT if I suppose that someone has a pain, then I am simply supposing that he has just the same as I have so often had."—That gets us no further. It is as if I were to say: "You surely know what 'It is 5 o'clock here' means; so you also know what 'It's 5 o'clock on the sun' means. It means simply that it is just the same time there as it is here when it is 5 o'clock."—The explanation by means of *identity* does not work here. For I know well enough that one can call 5 o'clock here and 5 o'clock there "the same time", but what I do not know is in what cases one is to speak of its being the same time here and there.
>
> (*Philosophical Investigations* I 350)

Wittgenstein is here appealing to some difficulty in understanding 'It's 5 o'clock on the sun' as evidence that one could not derive an understanding of 'He has a pain' directly, by a simple analogy, from a presumed prior understanding of 'I have a pain'. This move is still sometimes quoted as if it clearly established this point; but on closer examination the example will be seen to tell rather on the opposite side.

Why might we find the bare remark 'It's 5 o'clock on the sun' bewildering?

First, anyone with an adequate understanding of the 'o'clock' system of time reference knows that such references are incomplete and therefore indeterminate unless a time basis is either specified or assumed—e.g., 'by Greenwich Mean Time', 'by British Summer Time', 'by Eastern Standard Time'. The context usually sufficiently indicates such a basis, but the remark about the sun is presented without a context. We might reply, then: Supply whatever basis you wish and the meaning becomes clear. It is 5 o'clock on the sun by GMT just when it is 5 o'clock by GMT in Oxford or in Tokyo or anywhere else.

But, secondly, this reply presupposes distant simultaneity. If space-time is Newtonian, then there is such a thing absolutely, so there is no difficulty. If space-time is Einsteinian, distant simultaneity is not absolute but relative to this or that frame of reference. The obvious one to use is that in which the earth and the sun are almost stationary. With this frame taken for granted, 'It's 5 o'clock on the sun' is again determinate in meaning. Whether space-time is Newtonian or Einsteinian it will be 5 o'clock on the sun by GMT when some disturbance occurs there which is seen from earth (or observed as the cause of noise in a radio receiver) at 8 minutes past 5 by GMT.

Thirdly, place references are often used as a way of indicating a time basis: 'It's 5 o'clock in New York' may mean 'It's 5 o'clock by New York time'. The surface of the earth having been divided by convention into twenty-four time zones, 'It's 5 o'clock in X' can mean 'It's 5 o'clock by the time of the zone in which X lies'. Since 'on the sun' does not specify any place on the surface of the earth, 'It's 5 o'clock on the sun' cannot work in this way: it is for reasons thus made plain *not* analogous to 'It's 5 o'clock in New York'.

However, each time zone may be taken to include all places vertically above points on the earth which are within that zone. If so, 'on the sun' will specify whichever time zone the sun is vertically above at the moment. In this sense (and with the above-stated determination of distant simultaneity) it is never 5 o'clock on the sun: it is usually within half an hour of 12 noon by the time of the zone that the sun is vertically above, though it may be somewhat earlier or later where the zone boundaries wiggle a bit, or within half an hour of 1 p.m. where Summer Time is in force. In this sense 'It's 5 o'clock on the sun' is perfectly meaningful but always false.

Thus if we attend to what is already implicit in an adequate understanding of our ordinary uses of 'It's 5 o'clock', 'It's 5 o'clock in New York', and so on, we have no difficulty first in seeing why the bare contextless remark 'It's 5 o'clock on the sun' is initially bewildering, and secondly in seeing what—and how little—needs to be added to give it a clear meaning and a determinate truth-value. Having thus seen how 'It's 5 o'clock on the sun' is eminently understandable, we can return with increased confidence to the belief that if one had a prior understanding of 'I have a pain', one *could* derive from it an understanding of 'He has a pain', especially since there is here nothing analogous to the complications about a time basis, time zones specified by place references, or distant simultaneity.

University College, Oxford © J. L. MACKIE 1981

IS THIS A PROOF I SEE BEFORE ME?

By NEIL TENNANT

COLIN McGINN's paper 'Truth and Use' (M. Platts, ed., *Reference, Truth and Reality*, Routledge and Kegan Paul, 1980) presents a case against anti-realism of the kind espoused by Dummett. I believe the argument to be flawed in two respects. First he overstates the anti-realist view on what it is to grasp the meaning of a sentence. Secondly he is far too generous a radical interpreter in drawing realist conclusions from the imagined behaviour of some strange creatures he describes. I shall deal with these points in turn, but shall have to refrain here from exploring a theme suggested by some of McGinn's remarks—that of developing a suitably anti-reductionist account of meaning that could support realism. My purpose here is to be tersely critical rather than constructively expansive of McGinn's views.

The following is McGinn's statement of what he takes to be Dummett's view of what grasp of meaning consists in:

> Understanding a sentence . . . consists in an ability to determine its truth value in some canonical way, and acquiring a mastery of a language consists in acquiring such capacities with respect to its sentences. In fact, such a capacity is both executive and recognitional: it consists in a disposition, when prompted, to undertake a procedure which, in a finite time, terminates in a recognition, signalled by some overt gesture, that the sentence's truth condition is fulfilled or not fulfilled, as the case may be. It is because (actual) possession of such an effective decision procedure for truth value is a practical capacity that it provides the needed link between knowledge of sense and use (loc. cit. p. 21).

The misconception built into this straw account is the following. No intuitionist or anti-realist would demand that grasp of meaning should consist in an ability to decide truth value. What they maintain rather is that grasp of meaning consists in an ability to decide, *of any particular presentation*, whether it establishes the sentence as true or false. Here by "presentation" I mean something like a purported proof in mathematics; or an orderly series of pointings and indicatings in simpler ostensive situations; or a series of experiments, recorded observations and measurements in the natural sciences, documented and presented to the would-be decider so that he can check whether it establishes the result claimed.

McGinn's account has the following implication, which reduces it to absurdity. On his account, no intuitionist grasps the meaning of Goldbach's conjecture, which can be expressed in a not too complicated sentence of first order arithmetic. Now we do not know whether Goldbach's conjecture is true or false or undecidable on the basis of presently accepted axioms. We do not know whether we might not come across a proof of it in the future, or discover an isolated counterexample—that is, an even number greater than two that is not the sum of any two primes. We do not know whether a proof, if and when discovered, will require inexplicable ingenuity or will be seen, with the benefit of hindsight, as something that we must have been really dull not to have come up with before. All these things we do not know, and more. But this abyss of ignorance about Goldbach's conjecture does not—indeed cannot—divest it of determinate meaning for the intuitionist. Just because the intuitionist rejects the classical notion of truth—the notion of a sentence being determinately true or false, independently of our ability to recognize it as such—does not mean that he throws the Fregean baby out with that murky bathwater. The Fregean baby of course is the insight that we grasp meanings of sentences by grasping the meanings of their parts and also the way they are put together. *Of course* an intuitionist grasps the meaning of Goldbach's conjecture! For he grasps the meanings of its primitive parts and of the logical operators involved in its construction. His grasp of the former is displayed in his mastery of the recursive rules for plus and times. His grasp of the latter is displayed in his knowledge of what in general constitutes a proof of a sentence with the appropriate logical operator dominant.

We must distinguish between a sentence's having a determinate truth *value*, and having a determinate meaning, or truth *conditions*. McGinn is asking too much of the intuitionist in requiring that this very competent grasp, as just described, of the meaning of a sentence such as Goldbach's conjecture should underlie, or entail, or give rise to or inevitably be accompanied by an ability to evaluate it—that is, to get to work on it with an effective method that will terminate in a decision as to its truth value. This vitiates the following point in his case against Dummett:

> . . . a conception, encouraging assertion of bivalence, of a sector of reality that transcends . . . recognitional powers . . . could not be that in which . . . grasp of the sense of the relevant sentences consists; for, by Dummett's argument, nothing can qualify as *knowledge of meaning* unless it relates directly to some linguistic disposition of the speaker—a disposition, we have been agreeing, *to engage in a truth-value deciding procedure*—and the conception in question precisely does not relate to any such disposition (loc. cit., p. 29; my emphases).

McGinn is clearly demanding too much. The intuitionist may hit upon a proof after inspiration and creative work, the general lines of which may with hindsight be explained by appeal to his grasp of various aspects of the conclusion's meaning. Grasp of meaning no doubt somehow guides one's search for proofs, but it is only the proof *once found* that licenses an effectively defensible claim to *knowledge of truth value* (contrast with the quote above)—as I have argued in 'Language Games and

Intuitionism', *Synthese* 42, 1979, pp. 297–314. Grasp of the meaning does not guarantee a decision as to truth value, even on the intuitionistic account. As an intuitionist I could argue thus: 'I know what Goldbach's conjecture means because I know what it would be to have a proof of it before me. Don't ask me to tell you what that proof looks like! (That is, don't ask me for a, or the proof—for there may not be one.) To test whether I know, as I claim I do, what it would be to have a proof of Goldbach's conjecture before me, you could find one yourself and then have me check it; but that of course is not the only way.'

Indeed, we are not forced to ascertain grasp of meaning in this way only. Why can we not simply rely on the intuitionist's general understanding of his language, the language to which Goldbach's conjecture belongs? We have no right to insist that grasp of meaning be confirmed sentence by sentence, that we present for inspection proofs or disproofs of each and every sentence for which we raise the question whether meaning has been grasped. The intuitionist has compositional capacities like those of the classicist. He can understand individual words and operators and grasp sentential pedigree. And this allows him to grasp the meanings of new sentences, including Goldbach's conjecture. Moreover, the "basic grasps" involved can be ascertained by investigating his general ability to infer conclusions, reduce proofs to canonical proofs, find proofs of simple theorems, etc. etc.

In conclusion, then, I repeat: it is too much to require of the anti-realist that he display grasp of sentential meaning by successfully deciding the sentence. The anti-realist can admit the possibility of definitely meaningful yet undecidable sentences. McGinn does not attend closely enough to this distinction between being able to decide and being able to check that which purportedly decides.

I wish now to raise a further objection, this time against McGinn's "model case" in support of realism. He describes a thought experiment designed to show that realism is not only correct, but virtually unavoidable. Briefly, his imagined "community" of creatures is one of individuals rooted to the spot on the north side of a mountain. They cannot, "*in principle*", investigate the south side. Yet we are asked to suppose that they are possessed of a first order language with the operator S, taken to mean 'On the south side of the mountain it is the case that . . .' We are asked further to suppose that, 'like us, they are disposed, realistically, to assert bivalence across the board'. Nevertheless, continues McGinn, 'we have no account (from the anti-realist) of what it is about their use of language that confers upon their sentences the kind of sense a truth-conditions theory purports to record'. He then asks us to concede that 'the mountain-dwellers can manifest their knowledge of transcendent truth-conditions . . . by interpreting the assertions of fellow speakers as expressions of . . . realist beliefs'.

At first glance this sounds no different from the usual realist gestures at the widespread assertion of bivalence as part of the use of language that must be legitimated by any "meaning is use" theory. But the anti-realist's rejoinder here is well known. He considers rather some central feature of use on which to erect his theory, which can then enjoin peripheral reform. (Prawitz has made this point ably in his paper 'On the controversy between classical and intuitionistic logic', *Theoria*, 1979.) What is so different about the mountain-dwellers' case that they should be beyond this injunction, secure against any allegation of being philosophically muddled?

A great deal would seem to turn here on McGinn's claim (implicit in the case-description, but a claim that is nevertheless more in the realm of diagnosis than of listing symptoms) that the only or the best possible translation into English of the operator S in the language of the mountain-dwellers is the one given. On the one hand he appeals to their utter immobility as a limitation *in principle* on their capacity to apply decision procedures with respect to sentences containing S. But on the other, he appeals to their use of such sentences in order to attribute to them realist beliefs about what is the case on the south side of the mountain.

Now McGinn cannot have it both ways. There is a tension between these two aims, just as there is a tension between the two aims of finding a scientific theory that is as simple as possible, yet at the same time most strongly supported by the data (cf. Elliott Sober, *Simplicity*, Clarendon Press, 1975). I would argue that the case as McGinn describes it does not—at least, not without considerable expansion of pre-emptive detail that he does not provide—exclude translations of the operator S consonant with a non-realist account of the beliefs expressed by the sentences; beliefs that are now different, given the different translation of S, from those McGinn so readily attributes. The line of argument I am suggesting here is reminiscent of Hacking's in 'The Identity of Indiscernibles', *Journal of Philosophy*, 1976. Hacking argues that thought experiments about impoverished universes allegedly containing non-identical indiscernibles can always be countered by re-describing them, consistently with all the legitimate constraints imposed by their inventors, in such a way as to simplify their structure and make them conform to Leibniz's Law. No universe *has* to be described in such a way as to violate that law, and every universe *can* be described in such a way as to preserve it. (For example, Ayer's serial universe of sounds . . . ABCDABCD . . . can be re-described as the universe

A B
D C

with circular, clockwise (!) time.)

Now the same applies to McGinn's 'community' of rooted conjecturists. If we regard them as *in principle* incapable of south-side routines, what supports our attribution to the operator S of the meaning 'On the south side of the mountain it is the case that . . .'? The principle of charity can't license such a hand-out. McGinn's only supporting argument is '. . . a conception of that (for them) inaccessible reach of reality seems forced upon them if they are to have any reasonable explanation of the things they do observe: sheep disappearing and reappearing, etc.' How so? How can we, confronted by beings so crashingly non-paripatetic, attribute to them even a conceptual scheme of continuants? Given the constraints on their own movement, on re-orientation and measurement, on sizing up and sorting out, on separating, classifying, discriminating, manipulating, how can we even begin to attribute to them that conception of space similar to ours, against the background of which alone the thought fragment 'On the south side of the mountain it is the case that . . .' could be grasped?

If McGinn were to answer by expanding further on the capacities of his mountain-dwellers, one would expect the increased behavioural repertoire to sustain two concessions. First, it would underpin richer translations and richer attributions of belief to the mountain-dwellers, which would help McGinn. But secondly—and this would detract critically from the first concession—it would also undercut the alleged limitation on ability *in principle* to perform (in paradigmatic cases, as emphasized above) the kinds of procedure mastery of which underpins grasp of sense. The anti-realist insists that if we can show that a creature possesses certain beliefs by virtue, among other things, of grasping certain meanings, then, within the same realm of behaviour on the basis of which we show that, the creature must be able effectively to defend his assertions. The credit extended by charity should always earn this sort of interest.

An objector might at this point raise the problem of specifying precisely what it is about the abilities I have described the intuitionist as needing that distinguishes him from the classicist, or realist. For do not classicists also check proofs competently, and thereby display the same sort of ability? Certainly they do; but their proofs, which in general allow the classical moves of excluded middle, or double negation, or classical reductio or dilemma, to not provide them with recipes for the *effective* defence of their truth claims. The intuitionist, however, dealing as he does only with canonical proofs and effective procedures for reducing proofs to canonical form, *can* use his proofs in this way. *This* is the critical location of the emphasis on mastery of effective procedures, that I have been arguing is misplaced by McGinn.

University of Edinburgh © NEIL TENNANT 1981

REPLY TO TENNANT

By COLIN MCGINN

NEIL TENNANT (in 'Is this a proof I see before me?', ANALYSIS this issue) makes two criticisms of my paper 'Truth and Use' (Mark Platts, ed., *Reference, Truth and Reality,* Routledge and Kegan Paul, 1980). One criticism is that I incorrectly characterise Dummett's conception of what grasp of meaning consists in for an anti-realist, thus setting up a straw man for easy demolition; the other is that I fail to establish realism, either in general or for my model case of the mountain-dwellers. The first of these criticisms rests upon a total (and so far as I can see inexcusable) misunderstanding of what I wrote and of what Dummett argues; the second seriously misconstrues the drift of my paper, though the misconstrual is at least intelligible.

Tennant's mistake in his first criticism is to confuse what I reported as Dummett's suggestion, made on behalf of a *realist* about meaning, as to what grasp of truth conditions consists in, with—what is quite distinct—Dummett's own anti-realist conception of meaning. Dummett offers the realist an answer to the question how grasp of truth conditions is manifested—namely, by undertaking and completing a decision procedure—and then points out that the offered account will not extend to undecidable sentences; put differently, if the realist accepts the suggested answer he commits himself, absurdly, to the nonexistence of undecidable sentences. To avoid such consequences we must adopt an anti-realist semantics according to which sentence content does not exceed our actual limited powers of verification. So the conception of meaning Tennant says I pin on Dummett is precisely the conception he (Dummett) argues to be inadequate to the realist's needs, as indeed I expressly state in the sentence immediately following the first passage Tennant quotes. Nothing in my paper even remotely suggests that it is an objection to anti-realist semantics that it renders all sentences effectively decidable: that would be absurd. Rather, I was merely expounding Dummett's complaint that the realist can explain understanding only for decidable sentences, not raising the very same complaint against Dummett's *anti*-realist. Tennant's rejection of the latter complaint is then completely beside the point, though obviously correct; the view he criticises had never entered my head.

Tennant is also mistaken in supposing that I was out to establish realism. What I offered was a way of resisting an argument against realism; my conclusion was not that realism is, quite generally, obligatory —it was that it is permissible, at least so far as Dummett's arguments are concerned. Dummett argues that considerations relating to the acquisition and manifestation of linguistic competence *exclude* a realist theory of meaning. I contested this claim—but it is a further question whether realism, metaphysical or semantical, is actually true. My contention was only that Dummett's arguments do not show it to be false. I first argued that the acquisition considerations rest upon a naively empiricist doctrine of concept formation and theory construction, so that observation-transcendent realist conceptions are available to the mind. I then suggested that the manifestation argument be dismantled in two stages. First we should reject, as reductionist and behaviourist, the requirement that the content of sentences and thoughts be exhaustively explicable in terms both nonsemantic and behavioural. Secondly, we may, nevertheless, forge a *link* between understanding and use, by way of a theory of interpretation embedded in a theory of force. Analogously, if asked to explain how a chemist manifests his knowledge of chemistry, we can reasonably refuse to provide any behavourist *reduction* of his propositional attitudes, while agreeing that his chemical beliefs play a role in shaping the way he conducts himself in the laboratory. It is only to be expected that any content which transcends the observational will be thus indirectly tied to what a person does.

It may be objected against this that what a person does cannot exclude an anti-realist theory of content; and Tennant does so object to my treatment of the mountain-dwellers. That is very likely true, but not to the point. Dummett was not claiming, somewhat in the manner of Quine, that observable linguistic behaviour underdetermines sentence content, leaving it empirically unsettled whether a realist or an anti-realist semantics is correct. He was contending that the realist interpretation is *ruled out in principle* by the considerations about acquisition and manifestation. I undertook to show that realism is not to be *thus* ruled out. Perhaps I am partly to blame for Tennant's misconstrual, because I dropped some vague hints, *en passant*, about why realism in respect of the hidden side of the mountain would seem natural and even inescapable to the mountain-dwellers: they would need (e.g.) the idea of sheep persisting unobserved if they were to have any reasonable theoretical account of what they observe. But I did not intend to claim, and in fact explicitly denied, that I was undertaking a full-scale *defence* of realism. The model case seemed to me helpful because it enables us to disentangle Dummett's own anti-realist arguments from independent anti-realist considerations; for plainly *we* could not be led to question our own realist conception of the world on the strength of the theory of meaning demanded by *their* language. I was trying, among other things, to bring out the radical relativity of Dummett's arguments to the capacities speakers actually possess: if the capacities of the mountain-dwellers are not the measure of reality, why should *our* capacities enjoy this elevated metaphysical role?

I do not mean to imply that I still think 'Truth and Use' the best or most illuminating way of responding to Dummett's ideas, but I cannot see that Tennant's criticisms even engage with the line I was taking.

University College London © COLIN MCGINN 1981

IS EPISTEMIC PREFERABILITY TRANSITIVE?

By ROY A. SORENSEN

ALTHOUGH the 'heavier-than' relation is transitive, Edgington has shown that 'probably-heavier-than' is not ('On the Possibility of Rational "Inconsistent" Beliefs', *Mind*, October 1968).

> Suppose each of three boxes contains three objects which have the following weights, in pounds:
>
> Box A: 2, 6, 7
> Box B: 1, 5, 9
> Box C: 3, 4, 8
>
> Now consider repeated comparisons of two randomly selected objects, one from Box A and one from Box B. Since each of the three A objects is equally likely to be paired with each of the B objects, there are nine equally likely A–B comparisons. In five of the nine comparisons, the A object is heavier than the B object....

Edgington goes on to describe the B—C and A—C comparisons but I shall merely list the pairings:

A—B			B—C			A—C		
2, 1	6, 1	7, 1	1, 3	5, 3	9, 3	2, 3	6, 3	7, 3
2, 5	6, 5	7, 5	1, 4	5, 4	9, 4	2, 4	6, 4	7, 4
2, 9	6, 9	7, 9	1, 8	5, 8	9, 8	2, 8	6, 8	7, 8

In 5/9 of the possible outcomes the A object is heavier than the B object, so an object from A will probably be heavier than a B object. Similarly, an object from B will probably be heavier than a C object. Surprisingly, in 5/9 possible outcomes the C object is heavier than the A object, so an object from C will probably be heavier than an A object.

In 'A System of Epistemic Logic' (*Ratio*, December, 1972), Chisholm and Keim present "The Calculus of Epistemic Preferability". They take epistemic preferability to be a transitive relation but adopt as an axiom the 'broader principle according to which the relation of "not-being-epistemically-preferable-to" is transitive.'

$$(A2)\ [\sim(pPq)\ \&\ \sim(qPr)] \supset \sim(pPr)$$

One reads (A2) as: If it is not the case that p is epistemically preferable to q, and it is not the case that q is epistemically preferable to r, then it is not the case that p is epistemically preferable to r.

A counterexample to (A2) can be constructed if one takes Edgington's observation as a guide. Let A, B, and C be disjoint groups of equally talented researchers. Each of the groups has three subgroups whose sizes correspond to the weights in Edgington's three boxes. One subgroup from each of A′, B′, and C′ is randomly selected to work on three independent problems. The problems are equally difficult and are such that the likelihood of a correct solution is directly related to the number of researchers working on it. The three subgroups then submit their different conclusions, a, b, and c, respectively. By consulting the list of possible outcomes of this random selection, one can infer that a is more likely to be correct than b, and b is more likely to be correct than c, but surprisingly, c is more likely to be correct than a. Therefore, a is epistemically preferable to b, and b is epistemically preferable to c, but a is *not* epistemically preferable to c. So one has a counterexample to (A2).

Michigan State University

© Roy A. Sorensen 1981

ON AN ARGUMENT OF PEACOCKE'S ABOUT PHYSICALISM AND COUNTERFACTUALS

By N. M. L. NATHAN

THIS note criticises a new empirical argument for a 'token-token psychophysical identity theory', recently advanced by Christopher Peacocke in his excellent book *Holistic Explanation* (Clarendon Press, 1979).

> Suppose you touch a hot kettle and the pain causes you to withdraw your hand. Let this token pain event be $c\ \psi$, and let the token event that is the effect, the withdrawal of the hand, be e; and suppose that there is a certain event $c\ \phi$ that occurs in your nervous system at the same time as the pain event and that is a cause of the bodily movement e of withdrawal. About $c\ \phi$ we will make two empirical suppositions. The first is that we have a complete and wholly physical account of both the causal antecedents of $c\ \phi$ and also of the causal route from $c\ \phi$ to e in neurophysiological terms; this account completely explains how $c\ \phi$ causes e. The second supposition is that the event $c\ \phi$ is the only physical event occurring at the time it actually does that causes the bodily move-e (we are not excluding as causes nonphysical events occurring at that time) . . .
>
> Let us assume for *reductio* that $c\ \phi$ is not identical with $c\ \psi$. Then only two cases can arise: Either (1) $c\ \phi$ and $c\ \psi$ are jointly sufficient for e but do not overdetermine it, or else (2) e is overdetermined by $c\ \phi$ and $c\ \psi$. (134–5).

Peacocke excludes case (2) by pointing out that 'overdetermination implies that even if the pain event had not occurred, then withdrawal of the hand would still have occurred'. This, he says, 'we certainly ordinarily take to be false, and it is not clear why we should change the belief'. (135).

His argument to exclude case (1) is both more complicated and, as it seems to me, more vulnerable. He thinks that on the face of it (1) clashes with the assumption that we have a complete physical account of the causal route from $c\ \phi$ to e: 'If there were only joint sufficiency of $c\ \phi$ and $c\ \psi$ for e, then it would be true that if $c\ \psi$ had not occurred, then e would not have occurred, and so no such complete physical account would be available.' (135). He then concedes that such an account could still be available if $c\ \psi$ caused e only via its causation of $c\ \phi$, but argues that in this case it would not be true both that

(3) we possess a complete physical account of $c\ \phi$'s causal antecedents

and

(4) if $c\ \psi$ had not occurred e would not have occurred.

His reasoning here seems to be as follows. If $c\ \psi$ does cause $c\ \phi$, then (3) can be true only if there is an earlier physical event d which causes both $c\,\phi$ and $c\,\psi$. But in this case (4) will be false. To suppose that (4) is true in these circumstances is 'not to follow the way we assess the truth value of counterfactuals.' In assessing their truth value 'we do not allow, "that could only have been . . ." steps.' Suppose a man flips a switch and causes the light to go on a little time later. If we allowed 'that could only have been . . .' steps we would have to say, absurdly, that

(5) if he had not flipped the switch, the light would not have been on a little later

is false if the man in question is 'President of the local Electric Company, and the only way it could have come about that he did not flip the switch is by his having had to stay at his office to deal with a power crisis that would automatically have triggered emergency supplies and made all the lights in his own house go on'. (138).

It seems to me that Peacocke's Electric Company case does not in fact show that we should not accept counterfactuals on 'that could only have been . . .' grounds. This is because (5) does not adequately represent the counterfactual you are committed to by saying that the man's flipping the switch caused the light to go on, and the counterfactual you really are committed to is compatible with the counterfactual you would be committed to by reasoning that if the man had not flipped the switch that could only have been because he had to stay at his office, etc. If you say that the President of the company flipped the switch and caused the light to go on, then you must agree that if he had not flipped the switch a particular set of conditions S, jointly sufficient for the light to go on, would not have been realised. An indispensable member of that set would have been missing. But this is not to say that there is not some other set of conditions S′, of which his flipping the switch is not a member, which is also sufficient for the light to go on. And there is no incompatibility between 'If he had not flipped the switch, S would not have been realised' and 'If he had not flipped the switch, S′ would have been realised.' Suppose then that you accept (4) despite the fact that if $c\,\psi$ had not occurred, that could only have been because d did not occur, and then e would not have occurred either. Maybe this commits you to accepting, in the Electric Company case, that if the President had not flipped the switch the light would have gone on a little later. But even if it does, there is no problem, because this last counterfactual is compatible with any counterfactual you are committed to by saying that he caused the light to go on by flipping the switch.

University of East Anglia

© N. M. L. NATHAN 1981

'ALL ACTIONS OCCUR INSIDE THE BODY'

By E. J. LOWE

IN her recent book *Actions* (Routledge and Kegan Paul, 1980), Jennifer Hornsby develops a theory of action concerning which she admits that '[t]he result is a view about actions that is surprising on the face of it' (p. 1; all page references hereafter are to this book). Not least surprising is her avowal that 'I . . . claim that all actions occur inside the body. And that summarizes the principal thesis of this essay' (p. 14). My contention in this paper will be that Hornsby's 'surprising' view of actions arises from a serious misconception of the relationship between the notions of action and causation.

In order to understand Hornsby's argument for her 'principal thesis', it is necessary first to take note of her (admittedly quite useful) device of employing the subscripts 'T' and 'I' in order to distinguish *transitive* from *intransitive* readings of certain verbs like 'move', as illustrated in the sentences 'John moved$_T$ his foot' and 'John's foot moved$_I$'. (Other verbs like 'push' and 'pull' are clearly not ambiguous in this way.) The argument may then be summed up in Hornsby's own words as follows: 'Whatever events they are that cause the body to [move$_I$] they presumably occur inside the body . . . But movements$_T$ cause the body to move$_I$. And actions are movements$_T$. Thus . . . all actions . . . occur inside the body' (p. 13). (It should be remarked, incidentally, that the actual text contains 'move$_T$' where I have written '[move$_I$]': this correction seems necessary in order to make sense of Hornsby's' argument.)

Now there is one premise of this argument which I shall not question here, although not because I agree with it. This is that '[all] actions are movements$_T$', i.e. that all actions are to be *identified* with agents' movings$_T$ of their bodies (for instance, that, in the famous example discussed by G. E. M. Anscombe in her book *Intention*, Blackwell, 1957, the pumper's action of moving$_T$ his arm is *identifiable* with his action of poisoning the inhabitants of the house; cf. Hornsby, pp. 6 ff). But while I happen to believe that the criterion of action identity implicit here is mistaken, it doesn't appear to me that Hornsby's view would be any the *less* 'surprising' if it were restricted only to claiming that all *bodily movements*$_T$ (e.g. turnings$_T$ of eyes and noddings$_T$ of heads) 'occur *inside* the body' (my emphasis), so I shall limit myself to a critique of this narrower claim.

The crucial premise of Hornsby's argument is just this: that 'movements$_T$ cause the body to move$_I$' (p. 13). Clearly, since we have ample neurophysiological evidence that the causes of bodily movements$_I$ (such as turnings$_I$ of eyes and noddings$_I$ of heads) *are* events inside people's bodies—and more specifically events in their brains and nervous systems—then if those actions that are bodily movements$_T$ (such as turnings$_T$ of eyes and noddings$_T$ of heads) are *also* causes of the bodily movements$_I$, it seems that, short of postulating some quite implausible kind of causal overdetermination of these movements$_I$, we have no alternative but to identify the actions with the internal neural events. (My own view at this point is that Hornsby's suggestion that such an identity obtains is so counter-intuitive that her argument actually constitutes a *reductio ad absurdum* of the supposition that bodily movements$_T$ cause bodily movements$_I$; but, of course, to urge this here would merely be to beg the question against Hornsby).

Now let us just see, in a specific instance, what Hornsby's crucial premise commits us to saying. It commits us to saying that if John raises his arm, then his action of raising his arm *causes* his arm to rise. (The example is particularly convenient because English already supplies us with different forms for the transitive and intransitive senses of the movement verb involved.) But this, as I say, seems to me to betray a serious misconception of the relationship between the notions of action and causation. That the notions *are* intimately related is not in dispute: clearly, John's raising his arm is *somehow* connected with the fact that his arm is caused to rise. It is even clear enough that, if John raised his arm, then the counterfactual conditional 'If John had not raised his arm, his arm would not have risen' will very probably be true; though this in itself is certainly not a sufficient reason for saying that John's raising his arm *caused* his arm to rise.

We can, I believe, get some leverage on the problem if we allow our attention to be drawn to a fact that Hornsby herself notices, namely, that '[l]awyers believe that questions about whether persons have done certain things can be reduced to questions about whether they have caused certain other sorts of things, and numerous philosophers have spoken of actions as agents' *bringing things about*' (pp. 9—10; Hornsby's emphasis). (Obviously, the 'things' thus caused or brought about will be *events*.) Hornsby evidently believes that this manner of speaking is not only compatible with her views concerning the relations between action and causation, but actually tends to support them. But does it, really? According to this philosophico-legal manner of speaking, saying that John raised his arm is reducible to saying that John *caused his arm to rise* or *brought about the rising of his arm*. And this seems sensible enough, if a little clumsy from the point of view of English style. But does it, as Hornsby seems to think, give support to the thesis that *John's raising his arm* (i.e. his *action* of arm-raising) caused his arm to rise? I don't see that it does. Indeed, if John's raising his arm *is* a matter of his causing his arm to rise, Hornsby's thesis commits us to saying that *John's causing his arm to rise caused his arm to rise*. This is, to say the least, distinctly odd, and not, I would suggest, something to which the lawyers and philosophers in question would necessarily want to commit themselves.

All the same, to say that *A*'s causing *x* caused *x* (where *A* is an agent and *x* is an event, e.g. where *A* is John and *x* is the event of his arm rising) is not perhaps quite as manifestly absurd as to say, for instance, that *y*'s causing *x* caused *x* (where *both x and y* are events). After all, it may be urged, that *A* caused *x* is a matter of 'agent causation' whereas that *A*'s causing *x* caused *x* is a matter of 'event causation'. So perhaps a philosopher who maintained that there is an irreducible distinction between agent causation and event causation could without total absurdity, even if not without considerable implausibility, maintain that *A*'s causing *x* caused *x*. But Hornsby is surely *not* such a philosopher: she certainly *does* seem to believe that agent causation is somehow reducible to event causation. For she remarks: 'An implicit assumption of mine has been that "A person caused something" can be taken to say that an *event* in which the person played a role caused something' (p. 15, fn; Hornsby's emphasis). Now, it is not entirely clear whether in saying this Hornsby is endorsing the view that '*A* caused *x*' is reducible to some sentence of the form '*y* caused *x*', where *y* is an event in which *A* 'played a role' (whatever that may more precisely mean), or rather the view that '*A* caused *x*' is reducible to some sentence of the form '*y* caused *x* and *A* played a role in *y*' (her remarks in Chapter Seven, pp. 96 ff., tend to disfavour the former interpretation). But let us consider each possibility in turn.

First, then, suppose that Hornsby is endorsing the view that '*A* caused *x*' is reducible to some sentence of the form '*y* caused *x*'. Now in this case, if she *also* holds, as she apparently does, that '*A*'s causing *x* caused *x*' is true whenever '*A* caused *x*' is true (e.g. that if John caused his arm to rise, then John's causing his arm to rise caused his arm to rise), then she is evidently committed to what I have called the 'manifest absurdity' of maintaining that sometimes, viz whenever '*A* caused *x*' is true, a sentence of the form '*y*'s causing *x* caused *x*' is true. For if '*A* caused *x*' is reducible to '*y* caused *x*' this presumably implies that '*A*'s causing *x*' denotes the same event or occurrence as '*y*'s causing *x*', in which case '*A*'s causing *x* caused *x*' implies, by Leibniz's Law, '*y*'s causing *x* caused *x*'. And what is implied here surely *is* absurd. (Of course, someone who held that Leibniz's Law does not apply to, or has certain exceptions in the case of, causal statements, might attempt to avoid this implication. But it certainly doesn't appear that this is a line of defence that would sit at all comfortably with Hornsby's general position, particularly in view of her sympathy with a Davidsonian criterion of the identity of events in terms of the sameness of their causes and effects: see p. 135.)

Secondly, however, suppose that Hornsby is endorsing the view that 'A caused x' is reducible to some sentence of the form 'y caused x and A played a role in y'. In this case it cannot be argued that Hornsby is ever straightforwardly committed to maintaining the truth of any manifestly absurd sentence of the form 'y's causing x caused x', but still it appears, by parallel reasoning, that she is committed, by virtue of the thesis that 'A caused x' implies 'A's causing x caused x', to there being circumstances in which, for some x and y, y's causing x at least *partially* caused x (only partially now because an agent A's 'playing a role' in y will also have been a causal factor). But this certainly seems no *less* absurd a consequence.

From all this it should be clear enough that if, as Hornsby appears to believe, agent causation is reducible to event causation (and I personally have nothing to say one way or another on the matter), then, *pace* Hornsby, an agent's *action*, i.e. his causing something x (such as the rising of his arm) cannot without absurdity be regarded as a *cause* (in the event causation sense) of the event x (and notice that this is true independently of whether or not the action is merely a bodily movement$_{\text{T}}$). But if an action like John's raising his arm cannot then be regarded as a cause of his arm's rising, we can have after all no reason to suppose that the action occurs wholly 'inside the body', Indeed, the very question of *where* an action occurs would seem, according to this line of argument, often to be capable of no very *precise* answer at all. For if action sentences *are* ultimately reducible to causal sentences of the form 'y caused x' (or even only of the form 'y caused x and A played a role in y'), then to ask *where* (or indeed *when*) the action occurred is tantamount to (or at least involves) asking where (or when) y caused x, and this question itself may apparently have no very precise answer, especially if x and y are quite far apart in space (or time). (And on these grounds one might even question Hornsby's assumption, so far accepted uncritically, that actions are properly speaking *events* at all, though this is a large issue which I cannot go into here.)

University of Durham

© E. J. LOWE 1981

THE TRANSITIVITY OF CAUSATION

By GERMUND HESSLOW

I

IN a recent paper ('For want of a nail', ANALYSIS 40.1, January 1980, pp. 50—52) E.J. Lowe discusses the somewhat paradoxical fact that although causes are generally thought to be necessary and/or sufficient conditions for their effects, conditionship relations are transitive while causal relations are not. He illustrates this with the well-known nursery rhyme:

> For want of a nail the shoe was lost,
> For want of a shoe the horse was lost,
> For want of a horse the rider was lost,
> For want of a rider the battle was lost,
> For want of a battle the Kingdom was lost,
> And all for the want of a horseshoe nail.

Each line asserts a quite plausible causal relation, and yet the conclusion strikes us as absurd, or at least as somewhat peculiar. Both Lowe and J.L. Mackie in a comment on Lowe's paper ('The transitivity of counterfactuals and causation', *Analysis* 40.1, January 1980, pp. 53—54) approach the problem by considering various ways in which conditionship relations could fail to be transitive. According to Mackie, however, transitivity really holds, and the paradoxical impression made by the rhyme arises mainly because the conclusion, when considered in isolation, is rather surprising.

I find Mackie's suggestion unsatisfactory because, although I agree that the rhyme does not prove the non-transitivity of conditionship or of counterfactuals, I think that it does prove the non-transitivity of *explanations*. Even if it were true that, if the nail had not been lost, the kingdom would not have been lost, the loss of the nail is certainly not a very satisfactory explanation of the loss of the kingdom. The questions that should be asked concerning Lowe's example are therefore first, why some conditions do not explain, and second, why explanations are non-transitive.

II

An analysis of causation must proceed in two steps. Let us assume for the sake of argument that some definition of 'cause' in terms of necessary and/or sufficient conditions is essentially correct, and let us call an event so defined a *determining cause*. But such an analysis would be incomplete.

A barn has been burnt down, and it is claimed that the cause was a cigarette carelessly dropped in the hay. Under the circumstances, if the cigarette had not been dropped, the barn would not have caught fire. But this is also true of several other conditions such as the presence of inflammable material and of oxygen and the absence of dampness and an automatic sprinkler system. All of these factors can be regarded as determining causes, and I think that most of us would agree that they are, but we would also think it pretty silly if someone tried to explain the fire by pointing out the abundance of oxygen in the air. Still, a conditionship analysis will assign the same role to the oxygen as to the cigarette, something that becomes more evident if we change the example somewhat. Suppose that we have been living on a distant planet for some time, where the temperature is extremely high and where oxygen is absent. Now and then there is a leak in one of our oxygen tubes, and whenever this happens, inflammable objects in the vicinity burn. If asked to explain such a fire, we would probably say that it was caused by the oxygen. From a physical point of view the two situations are identical; nevertheless we feel that the cigarette in the first case explains the fire in a way that the oxygen does not. Clearly, the conditionship analysis must be supplemented with some selection criterion that will permit us to single out among the determining causes those that are also *explanatory*.

The difference between the two fires is, I think, this. When we try to explain why the barn caught fire, we are tacitly (and perhaps unconsciously) comparing this barn with other barns that did not catch fire, and we look among the determining causes for a condition that can *differentiate* this barn from the others. Oxygen and inflammable material fail on this account, for both factors, being equally present in all barns, fail to explain the difference between those barns that burned and those that did not. Only the cigarette succeeds. On our distant planet, on the other hand, heat and inflammable material are present everywhere, and only the variations in oxygen can account for the difference between the circumstances in which fires occur and those in which no fires occur.[1]

If this is correct, it introduces a certain relativity into causal explanations. What I am claiming is that an explanation of an event b generally involves a comparison between those circumstances c in which b occurred and a different set of circumstances c' in which b did not occur, but there will always be a number of different circumstances to which

---

[1] This talk about differences may seem to run counter to our intuitive conception of events, but it is established practice in many sciences. In genetics, for instance, it is the only way of separating the relative contributions of nature and nurture to the phenotype. It is impossible to say that the length of an individual is mainly determined by the one or the other, but it is possible to say that an individual's deviation from the normal is due to a genetic and not to an environmental difference.

c could be compared. The barn (considered as a set of circumstances), although normally compared to other barns, could also be compared to the dwelling house (where cigarettes are often dropped), in which case the presence of inflammable material would be part of the explanation. It could also be compared to the same barn before the fire, in which case the role of the cigarette as a precipitating factor would be essential.

A brief formulation of the criterion implicit in the above example is: an individual event *a* explains why *b* occurred in circumstances *c* while it did not in *c'*, iff *a* is a determining cause of *b* and events of kinds *a* and *b* do not occur in circumstances *c'*.

Thus, the cigarette (*a*) explains why there was a fire (*b*) in this barn (*c*), while there were no fires in those barns (*c'*), since the cigarette was a determining cause of the fire and neither cigarettes nor fires occur in the barns with which we compare. Oxygen however does, and so violates the criterion.

Let me add two brief comments. First, although the objects of comparison (or the circumstances associated with them) *c'* will normally be actual, this is not necessary. The burnt down barn can be compared to the neighbour's barn but also to a hypothetical 'typical' barn. Secondly the criterion is an idealization, which in practice will only be satisfied more or less approximately.

Now suppose that *a* explains why *b* occurred in *c* while *b* did not occur in *c'*, but not why *b* occurred *per se*[2], and that *b* explains why *d* occurred (in *e* but not in *e'*). Clearly, it does not follow that *a* explains *d*. *b* here functions as an ambiguous middle term. When explaining *b*, we concentrate on the difference between *c* and *c'*, but when we explain *d* (in *e* but not in *e'*) it is no longer relevant to compare with *c'*. An example will show how this could work. Smith's son blows up a balloon that bursts, and the sound gives Smith a start. The explanation of the balloon's bursting will depend on how we specify the question. What explains why it burst *now*, rather than a moment ago, is that it was inflated now. What explains why *this* balloon burst, when that one (which was also inflated) did not, is that this balloon was weaker. With a careless formulation we might say that the weakness of the balloon caused it to burst, and the burst caused Smith to start, but we could hardly say that Smith made a start because there was a weak balloon present. The central point here is the specification of the question and the reference state (i.e. those circumstances with which we compare). It is incorrect to say that the weakness explains why the balloon burst. It only explains the difference between this balloon and that one, and it was not this difference that gave Smith a start, but the bursting of a balloon at this particular moment. The weakness is an adequate explanation when we compare this balloon with that one, and both are inflated, but when we try to explain why Smith started, we have shifted our focus of comparison, and the inflation will not be a part of any circumstances with which we might plausibly compare Smith.

---

[2] To explain for instance the fire *per se* would, I think, mean accounting for the difference, not just between this barn and that one, but also between this barn and every other object. Presumably, all the determining causes would have to be cited.

III

Returning to our original example, we may note first that the loss of a nail does not satisfy our requirement as an explanation of the loss of a kingdom. Presumably, nails are lost all the time, even in the armies of non-lost kingdoms, and this cannot therefore account for the difference between this kingdom and the others. As in the balloon example, I think that the reason that transitivity fails here is that some of the links in the chain from nail to kingdom are ambiguous.

The loss of a rider does not really explain the lost battle. An army is always one man short of what it would have been, had there been one man more, but not all armies lose their battles. The lost rider can only explain the difference in outcome between this and a seemingly similar battle, in which an army had one man more and won. But in order to explain the loss of the kingdom, we need the lost battle *per se* and not just this difference.

Although some of them are less clear cut, similar remarks can be made about the other steps of the argument. As we follow the chain of determining causes, we successively change the focus of comparison with the effect that the explanatory relevance of the nail gradually disappears.[3]

University of Lund

© GERMUND HESSLOW 1981

---

[3] I would like to thank the editor and a referee of ANALYSIS, and also Bengt Hansson, for valuable comments on earlier versions of this paper.

THE IMPERSONAL FORMULATION OF THE COGITO

By Kim Davies

ONE line of objection to the Cartesian Cogito has contended that Descartes claims too much in saying 'I am thinking', and that all he is entitled to is an impersonal formulation such as 'thinking is going on'. I wish to counter an argument offered by Bernard Williams against this objection (*Descartes*, Penguin 1978, pp. 95–101).

The difference between Descartes' own and the impersonal formulation is that Descartes' account involves the existence of a thinker, while the impersonal formulation does not. Williams tries to show that the latter is untenable, on the following grounds. Suppose the following are true:

(T1) It is thought: P (T2) It is thought: Q

Will it follow that the following is true?

(T3) It is thought: P and Q.

No, says Williams, 'a distinct thought-content is involved in T3, and there is nothing in the occurrence of the two thought-events T1 and T2 to determine that that thought ever occurred at all. The thoughts T1 and T2 could be, as we might hopefully put it, "separate"' (p. 96). But in the possibility of such a separation lurks a difficulty for the impersonal formulation. For 'consider the following combination:

(T4) It is thought: it is not doubted whether Q
(T5) It is doubted: Q?

Is the thought reported at T4 true or false? Unless more is put in, nothing prevents its being straightforwardly made false, by the state of affairs T5. But granted what has just been said about T1, T2 and T3, it cannot be the case that the thought in T4 should have to be false just because of T5: we must want it to be possible that T5 be as "separate" from T4 as T2 can be from T1' (p.97).

What can we make of the parallel drawn here? If (T4) and (T5) are seen as instances of the schemata

It is thought: not P

and

P

then we have no parallel between the two pairs of sentences. However, if we reformulate (T5) thus:

(T5*) It is thought: perhaps not Q

which accords with the sense of the Cartesian doubt, then we get the following reformulation of (T4):

(T4*) It is thought: it is not thought: perhaps not Q

(T4*) and (T5*) are special cases of (T1) and (T2), and as a special case of (T3) we get:

(*) It is thought: perhaps not Q, and it is not thought: perhaps not Q.[1]

Now just as the truth of (T3) does not follow from that of (T1) and (T2), we should expect that the truth of (*) does not follow from that of (T4*) and (T5*), and this is surely so. Indeed, while (*) may be true, the thought which it reports is necessarily self-defeating, in that whenever it is thought, it is false. But while (T4*) and (T5*) are thus 'separate' in the way (T1) and (T2) are, this in no way prevents (T5*), and so (T5), from falsifying the thought reported at (T4*), and so (T4). For (T5*) is not 'separate' from (T4*) in another sense, in that it reports precisely that state of affairs held not to obtain by the thought reported at (T4*). This is not generally true of instances of (T1) and (T2), and is just what accounts for (T5*), and so (T5), making the thought reported at (T4*), and so (T4), false.

In the face of this, obvious enough, point, what sense are we now to make of Williams' insistence that (T4) and (T5) be as 'separable' as (T1) and (T2)? He goes on: 'T5 can falsify T4, we will want to say, only if the doubt-event T5 is not "separate" from the thought-event T4, or, one might say, if they both occur in the same thought-world (whatever that might turn out to mean)' (p.97). This is murky, but three points can be made. First: Williams here confuses (T4) with the thought reported at (T4). (T4) can be true while the thought which it reports is false. Thus (T5) can falsify the thought reported at (T4) without falsifying (T4) itself. This can only make it even less clear what Williams' point is, what problem he thinks we must deal with. Second: Williams has given us no good reason for claiming that (T5) can falsify the thought reported at (T4) only if (T4) and (T5) belong to the same thought-world. If it is doubted whether Q, in whatever thought-world, then the thought reported at (T4) is false. Of course, *if* we relativise the content of the thought reported at (T4) to a person, thus:

(T4R) It is thought: it is not doubted by A whether Q.

then that thought is not straightforwardly made false by (T5). But that hardly gives us a reason for making the relativisation in the first place. Third: on the impersonal formulation there is no basis for the individuation of, and so possible distinction between, thought-worlds. Given this, the thoughts reported at (T4) and (T5) cannot be 'separate' in the sense of occupying different thought-worlds, so that no problem about the falsification of the thought reported at (T4) (or (T4) itself) can arise from such 'separateness'.

---

[1] I am grateful to Jonathan Lowe for an illuminating discussion in which he suggested these reformulations.

Williams might be expected to accept this last point; but then, it might be objected, what becomes of our earlier acceptance that, on the impersonal formulation, the thoughts reported at (T1) and (T2) are 'separate'? Perhaps the idea is this: if the thoughts were not 'separate', in that they occupied the same thought-world, then the truth of (T3) would follow from that of (T1) and (T2). If on the impersonal formulation the thoughts are not 'separate', then this would require us to make the inference to (T3) from (T1) and (T2) which has already been deemed illegitimate, so that the impersonal formulation is incoherent. But this is hopeless. For, accepting the individuation of thought worlds via that of their thinkers which Williams is aiming at, it is notoriously the case that people can fail to draw even the simplest conclusions from their thoughts, and can indeed have thoughts which contradict each other. We cannot, then, get to (T3) from (T1) and (T2) even within the same thought-world. The sense of 'separate', such that the truth of (T3) does not follow from that of (T1) and (T2), has nothing to do with the sense of 'separate' as occupying different thought-worlds: so no incoherence is generated by the thoughts being 'separate' in one sense and not in the other.

Let us make a final attempt: a crucial part of the basis of the Cogito is the claim that, even while I doubt some proposition, Q, about the world, *that* Q is doubted is itself indubitable. It is the kind of *cogitatio* which secures the Cartesian anchor. If no doubt is possible here, then it looks as though (T4), which reports not merely a doubt whether Q is doubted, but a straight denial of its being doubted, must be ruled out. So if (T5) is true, the thought which (T4) reports cannot arise, and so (T4) cannot be true. But, the argument goes, we can make sense of the impossibility of such a thought arising only if the thoughts reported at (T4) and (T5) occupy the thought-world of one thinker; so that impersonal formulation, which rejects the postulation of such a thinker, cannot account for the incompatibility between (T5) and (T4), and so cannot account for the indubitability of *cogitationes* on which 'cogito' is founded.

While this reading has the merit of setting the issues more firmly in the context of Williams' discussion of the Cogito, it has nothing to do with (T1)—(T3), and is distinct from the question of (T5) falsifying the thought reported at (T4). Furthermore, the argument is unsuccessful. For the psychological impossibility that (T4) be true when (T5) is, is not the essential point. More important than psychological indubitability is whether there is any possibility of reasonable doubt of the existence of a *cogitatio* which is itself the object of immediate awareness.

Here the answer, on both formulations, is 'no'; *cogitationes* can be absolutely given as the objects of awareness in which no error is possible concerning their existence. This is all that is required for the impersonal formulation: any problems that arise about this foundational immediacy arise for both formulations alike.

On no interpretation, then, does Williams' decidedly obscure argument succeed in showing the incoherence of the impersonal formulation of the Cogito. It tends, rather, to confirm the impression that once we accept, however grudgingly (p.96), that formulation as meaningful independently of the existence of a thinker, there is no way out of the realm of pure *cogitationes*.

University of Durham

© Kim Davies 1981

DESCARTES, WELBOURNE, AND INDUBITABLE BELIEFS

By ROBERT FAHRNKOPF

MICHAEL WELBOURNE ('Cartesian Madness', ANALYSIS 40.1, January 1980) has proposed a novel interpretation of Descartes' *First Meditation*. This *Meditation* is usually regarded as having the purely negative function of systematically undermining all our beliefs, beginning with those based on the senses, in order to clear the stage for the introduction, in the *Second Meditation*, of the *cogito* as the one belief which, unlike all the others previously examined, is indubitable. Welbourne, however, thinks that Descartes is anxious even in the *First Meditation* to proclaim a number of indubitable beliefs; indeed, according to Welbourne, such a proclamation is the 'crucial point' of Descartes' discussion of madness.

This discussion is taken up when Descartes, after having suggested that we ought not to trust the senses because they sometimes mislead us, immediately counters this suggestion by giving examples of what I shall call optimal perceptual beliefs, such as that I am seated here by the fire, holding a piece of paper in my hands, and so on. After giving these examples, he adds: 'And how could I deny that these hands and this body are mine, unless I am to compare myself with certain lunatics . . . ?' (Lafleur translation, p. 76). And since Descartes does not regard the appeal to lunacy as a legitimate way to undermine these seemingly impregnable beliefs—presumably because it would be self-defeating for Descartes to offer to himself for *rational appraisal* as a basis for doubt the possibility that he is irrational—these beliefs continue to be unshaken.

Welbourne's surprising contention is that such beliefs remain unshaken throughout, and that Descartes regards these as examples of '. . . beliefs *about* himself or his body (not yet distinguished) which resist being called into question' (Welbourne, p. 48). But surely the whole *raison d'être* of the dreaming-strategy *is* to call into question, in a way which casts no aspersions on one's sanity, such optimal perceptual beliefs. Given that (a) Descartes *does* mean to challenge such beliefs as that I am sitting by the fire, etc., (b) the belief concerning my hands and body is cited in connection with these other optimal perceptual beliefs and seems to be treated on a par with them, and (c) in the beginning of the *Second Meditation*, before the *cogito* has been introduced, Descartes denies that he has as yet found something certain, we are entitled to presume, in the absence of compelling evidence to the contrary, that the belief that these hands and body are mine—which Welbourne sees as

Descartes' pre-eminent example of an indubitable belief—is no better off than the other beliefs undermined by the appeal to dreaming.

Welbourne's arguments for supposing that Descartes regards the belief concerning his hands and body to be pre-eminently indubitable are, first, that the denial that these hands and this body are *mine* is incoherent and, secondly, that Descartes himself acknowledges as much by admitting that he would have to be mad to doubt this. But Descartes says or implies no such thing; he says merely that he would have to be mad to allow the possibility that he might be mad to serve as a legitimate ground for doubting that these hands and body are his. He does not say that doubt of this belief on any basis whatsoever would be madness; in particular, the dreaming-strategy does provide an acceptable basis for doubt.

But, in any case, why does Welbourne himself suppose that doubt is incoherent here? If I understand him rightly—and the finer metaphysical nuances of his position may well have escaped me— he is saying that what makes this body mine is simply my consciousness of it as mine. That is, in becoming apprised of the existence of this body, there is nothing *further* I have to discover, and therefore nothing about which I could conceivably be mistaken, in order to determine whether this body is mine. Welbourne thinks he has textual evidence, in the *Sixth Meditation*, that Descartes shares these same subtle concerns about body-ownership. Welbourne cites in particular the passage about the pilot and the ship, but I cannot see that this passage—the point of which is to show that one's mind and body are more intimately related to one another than the pilot is to his ship—shows that Descartes takes Welbourne's position, or any particular position at all, on the question of how body-ownership is determined.

The *Sixth Meditation*, in fact, provides the best evidence *against* Welbourne's reading of Descartes. Near the beginning of that *Meditation*, Descartes reviews the history of his uncritical accumulation of beliefs based on the senses and his subsequent undermining of these beliefs. Included in his list of such beliefs is the belief

> . . . that this body, which by a certain particular privilege I called mine, belonged to me more properly and strictly than any other. For in fact I could never be separated from it, as I could be from other bodies (Lafleur, p. 130).

(This passage, incidentally, confirms Welbourne's view that in the *First Meditation* the belief in question is that this body is *mine*, rather than that this body *exists*, as the translation of Anscombe and Geach would have it. Welbourne supposes, wrongly, that I have tacitly followed the latter translation in my earlier article ['Cartesian Insanity', ANALYSIS 39.2, March 1979], but my purposes in that article are equally well served by either translation, since the belief in question is in any case one that Descartes proceeds to doubt.) Two paragraphs later, after having concluded his list of beliefs based on the senses, he continues: 'Later on, various experiences gradually destroyed all my faith in my senses . . .' (Lafleur, p. 130), whereupon he rehearses the familiar considerations, including the appeal to dreaming, already discussed in the *First Meditation*. That this review of beliefs based on the senses which were subsequently rendered doubtful specifically includes the belief about body-ownership amounts to virtually conclusive evidence that Welbourne has wrongly interpreted Descartes.

Still, if this belief is as obviously unassailable as Welbourne makes it out to be, how could Descartes have thought otherwise, as he evidently did? I suggest that insofar as we are willing to grant Descartes his dreaming-strategy to undermine perceptual beliefs generally, there is no additional difficulty involved in undermining the belief that this body is mine. If I am dreaming that I am a 350 lb. sumo wrestler, whereas the body I really possess, assuming I have one at all, is of a quite different size and shape, then it is obvious that my belief that *this* body, i.e., the body I seem to have in my dream, is mine is just false. Indeed, though perhaps less obviously, even if my dream-body exactly resembled my actual body, I would still be mistaken in supposing that the body to which I direct my attention in my dream is mine, because the dream-body does not actually exist.

Welbourne might still be right to maintain that, even of a dream-body which I take to be mine, I do not do so on any basis in addition to that which I have for believing that this body exists. But, of course, if I am dreaming, then my basis for supposing that this body *does* exist, and is thus mine, is radically deficient. At best, then, what is indubitable is the less ambitious belief that *if* I am not dreaming, then this body is mine. But if Descartes might be faulted for not recognizing the indubitability of this weaker claim, the proviso about dreaming nevertheless ensures, at least from the standpoint of the *First Meditation*, that the consequent of this hypothetical is itself *not* indubitable.

© ROBERT FAHRNKOPF 1981

MEMORIES AND THE WORLD

By D. H. M. BROOKS

IN his article 'Materialism and the Unity of Consciousness' (ANALYSIS 40.3, June 1980, pp. 144–6) Andrew Ward says that 'one cannot justifiably claim that the brain makes the unity of consciousness possible unless one's grounds for believing in the existence of matter do *not* rely on a best hypothesis argument' (p. 144). This is spelled out as the claim that it is unreasonable to accept both of two similar arguments commonly accepted by materialists. These arguments are for the existence of an external world of physical objects and for the reliability of memory. Ward's versions of the arguments are roughly as follows:

I. If there were no external world there would be no reason to expect the order of my perceptions to be as it actually is.
If there were an external world it would explain the order of my perceptions.
Hence, given the actual order of my perception, it is reasonable to believe that there is an external world.

II. If my past experiences do not connect with my present apparent memories (by leaving traces in a continuously existing brain which stores and reproduces their content) there is no reason to expect my apparent memories to match my past experience.
If there were such a connection it would be reasonable to expect my apparent memories to match my past experience.
Hence, given the match between apparent memories and past experiences, there is a connection between them.

Ward concludes that anyone who accepts both I and II is placed in a dilemma. He may either argue for I first and then for II or argue for II first and then for I. But he cannot first argue I and then argue II. This involves first accepting the external world on the basis of the actual order of perceptions and then later establishing that what one thought was the actual order is in fact guaranteed to be the actual order by the existence of the brain. Argument I presupposes the result of argument II because it relies on the actual order of perceptions which is only established by II. On the other hand one cannot proceed by first establishing II and then establishing I since the conclusion of II posits the existence of at least one material object, the brain. This means that the question is begged; matter is assumed to exist in the course of an argument to the existence of matter.

Such is Ward's claim so far as I understand it. In my opinion it is vitiated by at least two erroneous assumptions. Ward assumes that the argument for the existence of the external world must start from the *actual* order of perceptions rather than their apparent order and he assumes that one argument must be established before the other, not allowing that the two arguments can be established simultaneously. He also assumes that argumentation must only start from assumptions which are reasonable at the outset, not allowing that prima facie unreasonable assumptions may in the course of reasoning come to be established as facts.

In the rest of this paper I will try to show that these assumptions are erroneous and that one is quite justified in accepting versions of both I and II. First of all, the argument to the existence of the external world does not presuppose the accuracy of memory, since it does not have to start from the *actual* order of experience. This means that the link which Ward attempts to forge between the arguments does not hold. Ward takes pains to stress that the starting point is the actual order of experiences, but does not argue for this. Let us accordingly put ourselves in the position of the Cartesian doubter; at the present moment he has both present experiences and apparent memories. His apparent memories may or may not answer to past experiences of his. Insofar as the argument to the external world goes through, it is common cause that it goes through on the assumption that the apparent memories match past experiences. If this assumption is dropped we still have some explaining to do. We have to explain the order in which we seem to have experienced things, that we seem to be experiencing an external world and that we are conscious at all. The existence of an external world can be mooted to explain all of these things. It is not tied to a single explanandum, the actual order of sensations. It can be posited to explain the apparent order of sensations just as it is posited to explain the actual order. Moreover, it is generally used to explain both, and the degree of mismatch between them. If we see this as what the external world hypothesis explains, argument I can no longer be seen as presupposing argument II, as the link between the two arguments provided by emphasis on the actual order of experiences disappears.

A difficulty with the move from the actual order of experiences to the apparent order is that the positing of the external world is carried out on a basis which is scantier than it is commonly believed to be. Normally when one rehearses the simplest explanatory hypothesis argument for the external world one supposes that it both explains and is put forward to explain the whole order of sensations one has had in one's life up until the present. However, a virtue of Ward's argument is that it emphasises that an extreme sceptic will not grant one this, and will only allow one the present contents of consciousness. Contingent limitations upon the number of things one can be aware of at once mean that what one has in mind at present is far less than the totality of experiences one has already had by the time one is able to exercise one's faculties on philosophical problems. A sceptic might argue further that the contents of a single moment of consciousness are never numerous enough to justify the positing of a whole universe. A structure extending back through eons of geological time and outwards through the immense distances of astronomy is surely excessive as the explanation of a few momentary sensations! Might it not be simpler and more economical to suppose that a malicious demon had brought about this instant of consciousness?

If a single moment of consciousness is too meagre a basis for positing the external world, Ward may be correct in arguing that to refute the sceptic we have to assume the actual order of sensations. However, I contend that, when fully conscious, we are aware of far more than is commonly thought. There are borderline states between wakefulness and sleep when we are conscious of so little that the external world does, in fact, lose its grip upon us, so that if only we had enough of our wits about us, we might be serious sceptics. However, when completely awake, even though one may only be consciously attending to a single proposition, which is all one would mention if asked to list one's current thoughts, far more will, in fact, be present to consciousness. One generally has a pretty good subliminal awareness of one's surroundings which is evidenced by the speed with which changes and novelties are noticed. This awareness, moreover, is suffused with the past. Consider how one could, in looking along one's bookshelves, say of each of the books, how it was acquired, who its author is and what it contains. This knowledge though not self-consciously enunciated in looking along the shelf is constitutive of, amongst other things, one's present feeling of familiarity. Consider too how one's recognition of a thing depends on past experience of it. Present consciousness must also be reckoned as including the awareness of the immediate accessibility of a large store of memories of any number of different things. Present awareness is richer than we first think and much of this unnoticed richness points beyond the present moment. I conclude that the external world is a reasonable hypothesis to explain our present state of consciousness including what we currently believe to be the order of our experiences. Ward is wrong to suppose that the argument to the external world presupposes the actual order of experiences; we can posit the external world to explain the apparent order of experiences and our present state of consciousness, which is not too impoverished to support such an elaborate hypothesis.

Ward also assumes that the two arguments which he sees as presupposing each other have to be dealt with one after the other so that whichever we take first, it is vitiated by the fact that it presupposes the other. In fact, in replying to the sceptic, one can consider the arguments jointly. Let us consider the conclusions available to him set out in a table:

My apparent memories match actual past experiences	There is an external world
My apparent memories match actual past experiences	There is no external world
My apparent memories do not match actual past experiences	There is an external world
My apparent memories do not match actual past experiences	There is no external world

Each of these combinations can be adopted as a package deal and the first of them has the added bonus that its two components fit easily into a verified and comprehensive explanation scheme, which does not call for the positing of anything else. Veridical memories without an external world call for additional explanation of both the order of actual sensations and the veridicality of memories, while non-veridical memories with or without an external world demand an explanation of why they should seem to be memories when, in fact, they are not, and indeed why they exist at all. The case of non-veridical memories and an external world also requires an additional explanation. It is possible prima facie that we misremember everything, that at every present state we have veridical experiences which are forgotten and replaced by apparent memories of something different in the next moment. That we do not experience continual surprise or worse requires explanation in this case. The suggestion that we do, but keep forgetting it, is only plausible at moments of surprise. Moreover, since it is feasible that some form of evolution will take place in all worlds where there are candidates for it, it is unlikely that any being would survive unless there were a systematic mismatch of memories such that apparent memories of things worth avoiding and pursuing matched up with past experiences of things worth avoiding and pursuing. Further explanation of this mismatch mechanism must be given.

Even if it were the case that Ward's two arguments presupposed each other, this would not be vicious since the hypothesis formed by conjoining their conclusions is more plausible than the alternatives. There are a number of reasons for this. The gaps in the alternative hypotheses which require additional explanation have already been mentioned. There is also what might, somewhat barbarously, be called the 'elaborability' of the hypothesis. As we explore the consequences of positing an external world more and more satisfying explanations come to light and things come to hang together more and more. The malicious demon hypothesis is on the other hand minimally elaborable.

There is no explanation of how the demon causes us to have the awareness we do, there is, in fact, a radical discontinuity between transcendental and immanent causation. Nor is there any explanation of what distinguishes perceptions from apparent memories or why we think some mental items refer to the past. As we try to develop the hypothesis the transcendent nature of the demon leaves us in the position where we can say nothing justifiable.

The systematic elaborability of the external world hypothesis also justifies us in choosing it to explain a single present moment of consciousness. Once it is accepted, we can accept our memories as veridical and at once acquire a larger basis, as our past experiences become, in their turn, evidence for the hypothesis and part of what it has to explain. That the veridicality of memory is explained by the brain is itself further evidence for the external world hypothesis. Different aspects of our theory can be developed and checked against one another in the manner familiar from the sciences, where results from one science are used to develop or test another, and the effect of reducing one science to another is increased confirmation for both.

I conclude that it is not in any way objectionable to suppose that my present sensations and apparent memories are best explained by an external world containing my brain. Moreover, this explanation has the advantage that it provides intuitive and well developed accounts for phenomena which other hypotheses either leave unexplained or explain by wild and unfettered flights of fancy which may posit any of those weird members of the metaphysical menagerie such as malicious demons, Gods and heavenly clockwork.[1]

University of Cape Town © D. H. M. BROOKS 1981

---

[1] I would like to thank Denise Meyerson and Paul Taylor for their help with this paper.

ESCAPING THE SPEECH ACT DILEMMA

By KENT BACH

ESSENTIAL to Gricean theories of speaker meaning is the notion of reflexive (R-) intention. The main problem for such theories has been precisely how to characterize the requisite R-intention, in particular to give (1) a sufficiently strong account of its form and (2) a sufficiently weak account of its content. Attempts to produce (1) have resulted, in the face of successively more complex counterexamples, in successively more complex requirements on the R-intention a speaker must have if he is to mean anything in uttering something, but here our concern will be with (2), the problem of specifying the content of the speaker's R-intention that determines just what he means. This is the task of characterizing the sorts of R-intended effect a speaker (S) seeks to produce in an addressee (A), which effect determines what S means, but recently C. R. Carr (ANALYSIS 40.3, June 1980) has posed a dilemma for any such effort. He wonders whether it is possible to characterize the R-intended effect connected with speaker meaning in a way that, to avoid being too strong, does not turn out to be too weak.

Carr's specific target is a proposal by David Holdcroft in *Words and Deeds* (Oxford, 1978) who, borrowing an idea from Stephen Schiffer in *Meaning* (Oxford, 1972), suggests that not only the concept of speaker meaning but a taxonomy of illocutionary acts can be developed in terms of R-intended effects. Schiffer thought that the R-intended effect for statements is belief and for requests intention. But R-intending such effects is not necessary for making a statement or a request. Indeed, in stating S need not R-intend A even to believe that he (S) believes something; in requesting S need not R-intend A even to believe that S wants him to (intend to) do something. As John Searle pointed out long ago (*Speech Acts*, Cambridge, 1969), such effects are perlocutionary, not illocutionary.

> In the case of illocutionary acts we succeed in doing what we are trying to do by getting our audience to recognize what we are trying to do. But the 'effect' on the hearer is not a belief or a response, it consists simply in the hearer understanding the utterance of the speaker (Searle, p. 47).

Since understanding an utterance is more than understanding the sentence uttered, we must ask how a speaker, in performing an act of communication, R-intends his audience to take his utterance.

Aware that previous accounts are too strong, Holdcroft adopts the notion of 'providing a reason' (to think that S intends A . . .), but Carr finds this notion ambiguous and worries about its being interpreted too strongly. He cautions (p. 165) that 'There must be some sense of providing a reason other than actually attempting to get A to respond', since there are clear (though non-standard) cases of illocutionary acts in which S 'may clearly not intend to give any motivating reason for A at all'. And, Carr urges, some notion of providing a reason must be retained, for 'The reason being provided by the speaker very often does serve to individuate illocutionary force'. Thus, for example, the difference between ordering and begging someone to do something consists in the difference in 'reason for compliance, authority versus pity possibly, each provides' (p. 165).

How does Carr's dilemma arise? An adequate account of 'providing a reason' must allow for genuine but non-standard cases of, for example, ordering without intending compliance, so that ordering can be distinguished from begging even when compliance is not intended. But if ordering differs from begging solely in what S appeals to, S's authority rather than A's pity, and yet S does not intend (perhaps obviously) A to comply, then in what sense does S provide a reason for compliance at all? Not only must an adequate account of 'providing a reason' handle such non-standard cases, it must not allow acts like joking to count as illocutionary acts at all. For example, if A spills a drop of milk and S utters, 'I order you to shoot yourself', presumably S not merely does not intend compliance but is not giving an order at all. Such examples generate Carr's dilemma: an account of 'providing a reason' not too strong to exclude genuine but non-standard cases of illocutionary acts threatens to be too weak, making 'non-literal utterances literal' (p. 167; here I think that by '(non-)literal' Carr means '(non-)serious').

There is a real problem to be solved here but Carr's dilemma is illusory. It is illusory because there are two quite different roles that the notion of providing a reason plays in the theory of illocutionary acts. The dilemma can appear only if these two roles are not distinguished. To explain why I must introduce as background some central ideas in the Gricean account of illocutionary acts developed recently by myself and Robert M. Harnish in *Linguistic Communication and Speech Acts* (MIT, 1979). Here I cannot defend these ideas but offer them merely to show how Carr's dilemma can be escaped.

The basic idea is to define a notion of expressing an attitude that is distinct from the stronger notion of (R-) intending A to think one has it (stronger still is intending A to form a corresponding attitude). Performing a communicative illocutionary act (as opposed to a conventional one like christening or voting) consists simply in expressing an attitude. Its success consists in A's recognizing nothing more than the attitude expressed, irrespective of any attitudes that A may be intended to form or does form. So if S asks A to close the window, the illocutionary act of requesting succeeds if A recognizes the desire being expressed that A close the window. As a perlocutionary act the request can succeed only if A believes S to have that desire, intends to close the window, and does so. However, illocutionary (communicative) success requires only understanding—A's identifying the attitude expressed. In short, a request can be made and understood without being complied with.

Now the notion of providing a reason figures in the definition of expressing an attitude: 'For S to express an attitude is for S to R-intend the hearer to take S's utterance as reason to think S has that attitude' (Bach & Harnish, p. 15). By incorporating that notion this definition avoids requiring anything perlocutionary in the R-intended effect included in the content of an illocutionary intention, which is merely communicative. However, the reason a speaker provides by way of expressing an attitude is *not* the same reason that, as Carr observes, often serves to individuate illocutionary force, and so Carr's dilemma is really a mirage.

To see this consider again the difference between ordering and begging. It consists in the kind of reason connected with A's compliance. However, this is not the reason mentioned in the definition just given of what it is in general to express an attitude. Rather, this force-individuating reason is included in the content of the specific attitude expressed. So for S to order, rather than beg, A to do something is for S to express the intention that A take S's authority, rather than A's pity, as reason to comply, and that, applying the above definition, is for S to R-intend A to take S's utterance as *reason* to think S intends A to take S's authority (rather than A's pity) as *reason* to comply. Notice the double occurrence of 'reason': the first is essential for characterizing communicative illocutionary acts generally, whereas the second is what distinguishes particular kinds of 'reasons-identifiable' illocutionary act, e.g., ordering from begging.

Distinguishing the two points at which the notion of providing a reason enters into the theory of illocutionary acts dissolves Carr's dilemma. Nonetheless, there remains a problem not unlike the one he mentions, but it concerns only the first sort of reason, that provided in performing a communicative illocutionary act of any kind. The problem, as Carr recognizes, is to distinguish non-standard but genuine illocutionary acts, like ordering while obviously not intending compliance, from joking, which, as in the case of jokingly saying 'I order you to shoot yourself', is not an illocutionary act (of ordering) at all. Ironically, this specific problem is addressed by Bach and Harnish (pp. 57–59, 289–91), who focus on cases where S's insincerity is glaringly obvious. They argue that S can still express an attitude, such as intending A to do something, and thereby perform an illocutionary act, such as giving an order. Otherwise there would be nothing for S to be insincere about. Of course, when S is obviously insincere, he cannot rationally R-intend his utterance to be sufficient reason for A to take him as having the attitude expressed, but still he can R-intend his utterance as providing *a* reason, albeit overridden. In the case of joking ('Shoot yourself') there is no such R-intention. Rather than being obviously insincere in giving an order, S is not giving an order at all.[1] He does not R-intend his utterance to provide even a weak reason for thinking he intends A to shoot himself. Of course it could be so taken, indeed be taken as intended to be so taken, but S has no R-intention that A so take it. Thus, as Bach and Harnish argue, S is not ordering A to shoot himself. S said 'Shoot yourself' to A but was not being serious. If he was giving a real order, then if he did not have the intention he was expressing (that A shoot himself), he would be insincere rather than unserious.

Both of the above cases, as well as nonliteral utterances, can be described as, 'S said something but did not mean it', a locution whose ambiguity can yield confusion. Is S being described as speaking (a) unseriously, (b) insincerely, or (c) nonliterally? Joking is a case of (a), Carr's 'non-standard' ordering an example of being insincere (b). The difference is between not expressing an attitude and not having the attitude one is expressing. As for (c), one could speak seriously and sincerely but not mean what one says in the sense of not speaking literally. If one utters 'I could never survive another of his lectures', the belief expressed is likely to be other than the belief that another such lecture would be fatal. That is the attitude one would be expressing if speaking literally. In speaking nonliterally S R-intends A to take his utterance as reason to think that S has a certain attitude other than the one he would be expressing if he meant what his words meant. (See Bach & Harnish, ch. 4, for a detailed account of the patterns of inference followed by the addressee to identify the attitudes that a speaker expresses in performing an illocutionary act nonliterally.)

In sum, the notion of providing a reason plays a dual role in the theory of illocutionary acts. It serves both in characterizing illocutionary acts generally in terms of expressing an attitude and in distinguishing certain specific kinds of illocutionary acts from one another. Understanding the first role enables us to distinguish serious from unserious utterances, sincere from insincere utterances, and literal from nonliteral utterances, and so long as this role is not confused with the second, Carr's dilemma is avoided.

San Francisco State University

© KENT BACH 1981

---

[1] Holdcroft's theory, Carr's primary target, seems unable to accommodate the difference between these two cases. In his own reply to Carr (ANALYSIS 40.4, 1980, p. 210) he claims that in cases of joking his crucial condition (iv) is not satisfied because the speaker 'only pretends to satisfy (iv)'. Since the speaker makes it clear that he is joking, the hearer is provided with 'an overriding reason for not acting'. However, there remains the question whether the speaker intends to provide the hearer with any reason at all. Accordingly, an adequate formulation of Holdcroft's condition (iv) must take into account the difference between obvious insincerity and unseriousness. To do so it would, I believe, have to incorporate the notion of expressing an attitude.

DIVINE DETERMINATENESS AND THE FREE WILL DEFENCE

By DAVID L. PAULSEN

IT is the thesis of this paper that one cannot consistently hold that

(i) God has an eternally determinate nature; and
(ii) the free will defence is an adequate response to the "problem of evil".

My strategy, in defending this thesis, will include

(a) a brief sketch of the free will defence to the problem of evil,
(b) a review of a modern rejoinder to the free will defence; and
(c) an analysis of two important responses to that rejoinder; and, finally,
(d) an argument that these responses are fundamentally incoherent.

The so-called "free will defence" has long served as a bulwark against the claim that given the evil we experience in the world, one believes in Christian theism only on pain of rational inconsistency. God could have prevented the evil we find in the world only by depriving men of their freedom, but, being a morally perfect Being, God would have created the best possible world, and a world consisting of persons who freely choose either good or evil is better than a world where persons, deprived of freedom, are compelled to choose only good.

In response, critics of the free will defence have claimed that its defenders present us with a false (i.e. non-exhaustive) dichotomy. For God was not limited to choosing either World I—a world with freedom and, consequently, moral evil—or World II—a world with no freedom and no moral evil. A much better alternative was available: God, being omnipotent, could have created World III, consisting of persons who would always freely choose the right; that is, God could have created a world with both freedom and no moral evil. The criticism presupposes the compatibility of determinism (soft) and freedom. Freedom is here understood as the absence of external constraint; on this view, I am free if there is nothing external to me preventing me from choosing that which I desire to choose or compelling me to choose that which I do not desire to choose. God could have thus brought World III into existence by (1) creating persons with characters so righteous that they only (or always most strongly) desired that which is right, and (2) insuring that nothing external to persons prevented them from choosing and acting in accordance with those desires.

Anticipating and/or responding to this line of critique, John Hick (*Evil and the God of Love*, Harper & Row 1966) and F. R. Tennant (*Philosophical Theology*, vol. 2, Cambridge University Press 1930), in two of the more impressive and closely reasoned modern theodicies, have argued for a stronger notion of freedom than the soft-deterministic version employed by critics of the free-will defence. They affirm that God endowed men with (what some call) contra-causal or categorical freedom—the creative power to mold and modify their own characters.

Hick writes:

> Is there a third concept of freedom such that a free act is neither, on the one hand, the inevitable outworking of a man's character nor, on the other hand, a merely random occurrence? There is such a concept of freedom, and indeed, it is the one that seems intuitively most adequate to our ordinary experience as moral agents... It is roughly the notion of freedom as limited creativity. This must be thought of as involving an element of unpredictability; for, whilst the action proceeds from the nature of the agent, the nature from which it proceeds is that of "the actual self alive in the moment of decision" [citing Hartshorne's *Logic of Perfection*]. Thus, whilst a free action arises out of the agent's character it does not arise in a fully determined and predictable way. It is largely but not fully prefigured in the previous state of the agent. For the character is itself partially formed and sometimes partially re-formed in the very moment of free decision (p. 312).

Tennant's notion of freedom is similar to Hick's. He states that the free will defence is:

> sometimes alleged to be worthless, on the ground that in creating human wills God created all their volitional issues. But this objection involves the crude notion that a will is a concrete thing, like a machine instead of a name for an activity ... Although ... God assigned to each soul its specific nature or individuality, determinative of its original response (when it should become embodied) to primary presentations, and prescribed the scope of certain of its capacities and faculties, this was not to foreordain as a chain of quasi-mechanical effects, all its future actions, and so to be responsible for its misdeeds (pp. 189–90).

In explaining why God would consistently prefer World I over World III, Hick argues that moral virtue acquired through a process of free and self-directing struggle and development is more valuable than virtue created *ab initio*. He says:

> ... (T)he value judgement that is implicitly being invoked here is that one who has *attained* to goodness by meeting and eventually mastering temptations, and thus by rightly making responsible choices in concrete situations, is good in a richer and more valuable sense than would be one created *ab initio* in a state either of innocence or of virtue. In the former case ... the individual's goodness has within it the strength of temptations overcome, a stability based upon an accumulation of right choices, and a positive and responsible character that comes from the investment of costly personal effort (pp. 291–2, emphasis added).

Tennant makes a stronger claim. Unlike Hick who implies that moral goodness might be had ready-made, albeit of inferior grade, Tennant argues that moral goodness is impossible unless it is acquired through a *developmental* process guided by free decisions. He asserts:

> And if the world is to realize the good, that being its *raison d'etre*, it would seem that it then must be characterized by development or epigenesis. For moral goodness cannot be created as such; it cannot be implanted in any moral agent by an 'almighty' other. It is the outcome of freedom, and has to be acquired or achieved by creatures (p. 185)

Tennant continues:

> But the best possible world implies the existence of moral agents; its crown cannot be the puppet or the automaton. Were our conduct determined like the movements of the machinery of a clock, our world might manifest a pre-established harmony and fulfill the purpose of the clock-maker . . . (*But*) (*t*)*here is no moral goodness in a clock, however perfectly it may keep time* (pp. 188–9, emphasis added).

By way of reply, then, to critics who claim God could have created World III—a world consisting of beings who always freely choose the right—

(i) John Hick implies that such a world would be possible, but would be morally *less valuable* than World I; and
(ii) F.R. Tennant asserts that World III would be without any moral value whatever.

These responses appear viable until one considers that Tennant, explicitly, and Hick, implicitly, affirm that God has, eternally and immutably, a morally determinate nature.

Tennant says:

> God without a world is a superfluous abstraction, and a God who might have 'chosen' a different seminal world from this, or different 'primary collocations', would be a different God . . . Determinateness excludes what are called possibilities, or other subsequently conceivable modes of action; choice between them, as if none were ruled out by the determinate nature of the Deity, is thus a superfluous supposition. The world is what it is because God is what He is. (pp. 183–4).

Again Tennant says:

> In that God is love, He is not hate; in that He wills a developing moral order He is not the creator of a paradise of angels. Possibilities are thus determined by what God is; and God is limited by His very determinateness, not an indeterminate Absolute in whom all differences are lost (p. 188).

The basic incoherency in Tennant's position can now be identified, if it is not obvious already. He asserts that:

(1) Moral goodness cannot be had *ab initio*. Rather it must be the outcome of a self-directed *developmental* process. 'There is no moral goodness in a clock, however perfectly it may keep time.'

(2) God has a totally and eternally determinate nature, given which, only the *actual* issues of that nature are *possible*—i.e., God *can* do and choose only what He does do and choose. The possible and the actual are co-extensive.

(3) God is morally good; indeed, morally perfect.

But given the assertions (1) and (2), God cannot be morally good because His character is *not* the outcome of a self-directed developmental process. Accordingly, Tennant's position is logically incoherent. Either he must concede that God could have created morally good (perfect) persons *ab initio* (which he explicitly rejects) or he must concede that God is not morally good, no matter how perfectly He acts.

The dilemma could be avoided by asserting that if God has a determinately good character now, the same has been forged in aeons past, by a means of self-directed developmental process, but this idea doesn't fit with Tennant's concept of the nature of God.

John Hick's view falls prey to the same kind of incoherency. While implying that virtue could conceivably be had *ab initio*, Hick nonetheless asserts that virtue *acquired* or *developed* through a self-directing process is "richer" and "more valuable" than virtue had *ab initio*. But presumably on Hick's view, God has *not acquired* or attained His virtue through a self-directing developmental process. Accordingly, Hick's view seemingly entails that men may achieve a kind of virtue which is richer or more valuable than God's virtue—a conclusion which Hick would undoubtedly reject.

The dilemma appears final. Either Hick's and Tennant's responses to the critique of the free will defence must be given up or they must modify their conceptions of God. Rational consistency precludes their having it both ways.

Brigham Young University

© DAVID L. PAULSEN 1981

CARR ON BELIEFS AND DISPOSITIONS

By PETER SMITH

IN ANALYSIS 40.4, October 1980, David Carr suggests that the familiar fact that belief ascriptions are (in a sense) negatable in two ways indicates 'that something more than a simple dispositional property is expressed by the term "belief"'. He goes on to argue that ascriptions of intentions to act exhibit a further complexity, being negatable in three ways, and so again 'appear to assert something that defies any resort to talk of dispositions, either single or multi-track'. Appearances here are evidently deceptive; for the logical features which Carr notes are no bar at all to dispositional theories of belief or intention.

Carr does not say exactly what he treats as a dispositional property. But I take it that, if the philosophical tradition is to count for anything here, it will be a sufficient condition of the predicate 'ϕ' expressing a dispositional property that 'X is ϕ' is logically equivalent to 'X is such that, if X were in circumstances C, X would (do) ψ' for some fillings for 'C' and 'ψ'. So, the claim that beliefs cannot be dispositions because belief ascriptions can be negated in two ways entails the following: 'X believes that p' could never be equivalent to a sentence of the form 'X is such that, if he were in circumstances C, he would do ψ', *because* the former sentence is negatable in two ways. But this implication of Carr's claim has only to be stated to be seen to be false. Consider, for example, the old style of quotational analysis of belief, according to which 'X believes that snow is white' is equivalent to something of the form 'X is such that, if in circumstances . . . , he would assertively utter (some sentence equivalent to) "snow is white"'. Whatever is wrong with this as an analysis, it can't be faulted on the ground of failure to reflect the fact that the original belief ascription can be negated in two ways. On the contrary, corresponding to the difference between 'it is not the case that X believes that snow is white' and 'X believes that snow is not white' we have the patently different results of negating the whole analysans, and inserting a 'not' in the quoted occurrence of 'snow is white'. Indeed, a proponent of the quotational analysis might claim it as a virtue of his account that it explained just how it can be that the original belief sentence *is* negatable in two significantly distinct ways: but for all that, the quotational analysis treats beliefs as dispositions.

Similar remarks apply mutatis mutandis to the case of intentions. Consider, for example, the suggestion that 'X intends to bring it about that p' is equivalent to something of the form 'X is such that, if in circumstances . . . , he would assertively utter (some sentence equivalent to) "I will bring it about that p"'. Whatever its faults, the analysans at least reflects the fact that the original sentence is 'negatable in three ways'; for evidently we may negate the whole analysans, or insert a 'not' in two ways into the quoted sentence it contains. So again the logical feature which Carr emphasizes *can* be captured quite smoothly by at least one type of dispositional theory, and thus the idea that the logical feature in question should quite generally lead us to 'view attempted dispositional analyses . . . with incredulity' has no warrant at all.

University College of Wales, Aberystwyth © PETER SMITH 1981

KILLING PEOPLE INTENTIONALLY, BY CHANCE

By Kim Davies

IN his discussion 'Peacocke and Kraemer on Butler's Problem' (Analysis 40.3, June 1980), E. J. Lowe argues that the symmetry between the cases involved in Butler's problem is such as to render our intuitive ascriptions of intentionality paradoxical. On the basis of further investigation of the symmetry, I offer a justification of our intuitions and so a solution to the problem. Look at these two situations. One is an ordinary dice game, won on the throw of a six. The second is a version of Russian Roulette, called RR2, in which the player spins the barrel of a six-chambered gun containing one live cartridge, aims at the opponent, and pulls the trigger. If the bullet is fired, and hits the opponent, the game is won. We suppose that the death of the opponent is a causal consequence; to make the parallel exact, let us suppose that as a causal consequence of throwing a six in the dice game the Mona Lisa is presented to the thrower.

Now look at an asymmetry between throwing a six and killing Smith. If there is evidence that Brown intended to kill Smith anyway, not just in the game of RR2 but by any suitable method; and if the game of RR2 was just one suitable opportunity (perhaps because it helped to disguise his intentions), then we would call the killing intentional. However, we would have to say that it was not intentional that *on this particular occasion* Brown kill Smith. For, having pulled the trigger, he has no control over whether the bullet is fired, nor over killing Smith, and it is inappropriate to describe doing something as intentional in the absence of any control over whether it is done. Similarly, if Brown intends, one way or another, to get the Mona Lisa, and the dice game is just one opportunity, then we would say that he got it intentionally, but not that he threw a six intentionally.

The asymmetry between throwing a six and killing Smith hangs on the fact that killing Smith is a consequence of hitting Smith with the bullet; a consequence which, unlike throwing a six, is the object of a general intention. That is, the scope of this intention is not limited to a particular occasion. While we can say that, on this particular occasion (of an RR2 game), it was intentional that Smith be killed (thus distinguishing it from other possible occasions where Smith is killed unintentionally), we cannot say that it was intentional that, on this particular occasion (of an RR2 game), Smith be killed. Similarly, getting the Mona Lisa, a consequence of throwing a six, can be the object of a general intention without it being intentional that in this particular attempt it be obtained. Neither the firing of the bullet under the conditions of RR2 nor the throwing of the six are intentional, since, having decided to make the attempt, Brown has control over neither. This case seems to me to be the basis of our intuitions here, and to show why we would say that Smith's killing, and not the throwing of the six, was intentional. Several important points remain to be made.

First, if it is the general intention to kill Smith, not limited to the conditions of RR2, which allows us to call the killing intentional, could we not have a matching intention to throw a six, not limited to the conditions of dice games? The answer is yes. That is, if we were prepared to replace the ordinary die with a cube having a six on each face, and if throwing a six with such a die would satisfy our intention, then we could say that, in the ordinary dice game, Brown intentionally threw a six (i.e., it was an object of a general intention), though it was not intentional that on this occasion he throw a six. The reason that calling the throw of a six intentional is so counter-intuitive, is that in talking of wanting to throw a six we presuppose that the normal conditions of die-throwing will obtain, i.e., that it is on any particular occasion beyond the control of the thrower and that if it were not beyond the player's control the six would not count. When talking of wanting to kill someone, we do not presuppose that the conditions of RR2 will obtain, i.e., that on any particular occasion it is beyond the control of the thrower. For we count someone as having killed someone else, whether it was in a game of RR2 or not.

Second: does the killing of Smith have to be the object of a general intention, which would consider and accept suitable alternative means to its end, in order to be intentional? Could Brown not set up the RR2 situation with an intention to kill Smith in this particular game of RR2, and with no interest in Smith's death by other means or at other times? I think that, while Brown would rightly be held *responsible* for Smith's death, the answer to the second question must be no. Brown may want to kill Smith under these conditions, but a want, without control over the desired outcome, cannot itself constitute an intention. All we can properly ascribe to him is the intention to have a shot at it, so to speak. Given *this* intention, the ascription of intentionality to the actual killing implies a control over the killing which, having pulled the trigger, Brown does not have. However, since this is such a special occurrence, the fact that we should not call the killing intentional need not be thought counter-intuitive. Indeed our intuition might be that, if Brown genuinely intended to kill Smith he would have shot him there and then, without any of this barrel-spinning business.

Third: in the standard case described above, while the killing of Smith was intentional, it seems that hitting him with the bullet (the equivalent of throwing a six on the left-hand side of the table) was not intentional—even though hitting Smith with the bullet is surely what we mean by shooting Smith, and out intuition is to say that the shooting, as well as the killing, of Smith was intentional. Meanwhile our intuition about the parallel act of throwing a six is that it was not intentional. Does the paradox lie here? The problem here can be solved by again looking at the normal conditions of throwing a six—where a one-in-six chance is central—and of shooting people. Someone counts as shot whether it was a one-in-six chance or not. Normally it is not, normally it is within the agent's control, and so normally it is intentional that, on the particular occasion, someone be shot. To make the parallel with throwing a six exact, we should describe the case as one of 'shooting Smith under the conditions of RR2', or 'hitting Smith with the bullet in a game of RR2', and not as 'shooting Smith' simpliciter. It does not seem counterintuitive to say that the shooting of Smith under the conditions of RR2 is not intentional, for, having entered the game, Brown has no control over its result.

Fourth: what part does hope play in all this? For Brown's hoping to kill Smith as he pulls the trigger seems an important factor in our intuitive ascription of intentionality to the killing. While some people may kill themselves intentionally by playing Russian Roulette, others may have no desire whatever to die, and we do not intuitively say that they kill themselves intentionally. It is just bad luck. The same goes, mutatis mutandis, for Brown and Smith. The hope figures as evidence of Brown's general intention to kill Smith. There may be cases where it is insufficient evidence. If RR2 is set up as a test of nerve and Brown, to his surprise, finds himself hoping that he does kill Smith as he pulls the trigger, and if he then fails, and heaves a sigh of relief, we do not normally count this momentary hope as evidence of intention. If, on the other hand, it is followed by future attempts on Smith's life, we might then cite this hope as the emergence of an intention to kill Smith. Of course, if Brown succeeds, we have no way of deciding this issue, which may be the cause of some guilt feelings on Brown's part. We would have to settle counterfactual questions such as whether, if Brown had lost all interest in the game for its own sake, he would still have pulled the trigger, or whether, if he had failed, he would have sought other opportunities to kill Smith. What is at stake here is determining whether Brown did act under a general intention to kill Smith, and the mere presence of hope, of the wish that he kill Smith, is not usually sufficient. Where the intention *is* there, we have the standard situation discussed earlier.

Finally, consider Lowe's most striking presentation of the paradox, in which Brown throws a die constructed as a bomb which will explode if and only if it turns up six, and in which Brown throws the die hoping that it explodes, and knowing that if and only if it explodes it will kill Smith. Lowe suspects that we would say that Brown exploded the bomb and killed Smith intentionally, but not that he threw a six intentionally. I suspect he is right, but disagree that it would be unjustified. For if Brown's hope that Smith die is evidence that Brown is acting under a general intention to kill Smith, then he kills Smith intentionally. Further, either exploding the bomb should be redescribed as 'exploding the bomb under conditions like those of RR2', in which case it is not intentional; or we should take Brown as intending generally to explode a bomb killing Smith, though not necessarily under these conditions, in which case Brown explodes the bomb intentionally, though it is not intentional that at *this particular attempt* he explode the bomb, since having embarked on the attempt he has no control over its outcome. Throwing a six is not intentional, for familiar reasons. What we cannot allow, although the focus on the particular situation seems to lead us to it, is to say that it is intentional that on this particular occasion, under the conditions of RR2, Brown kill Smith; for having embarked on the attempt Brown has no control over whether he throws a six, explodes the bomb, or kills Smith. Our intuitive ascriptions of intentionality sensibly take into consideration intentions which are not limited to particular occurrences in this way.

University of Durham

© KIM DAVIES 1981

KILLING, LETTING DIE, AND EUTHANASIA: A REPLY TO HOLLY SMITH GOLDMAN

By PHILIPPA FOOT

GOLDMAN tries to show that 'so far as the intrinsic nature of the act goes, we cannot accept passive euthanasia without accepting active euthanasia as well' (ANALYSIS 40.4, October 1980, p. 224). Her argument runs as follows.

If it is true that, other things being equal, it is worse to kill someone than to allow him to die, and worse to set his house on fire than to fail to extinguish a fire that is burning, *it is true only because these are applications of the general principle that, other things being equal, it is worse to harm a person than merely to allow this harm to come upon him.* She concludes that as an act of euthanasia is acceptable only when death counts as a *good* to the person who dies, the principle does not apply, and therefore the distinction between active and passive cannot affect the acceptability of acts of euthanasia.

I do not know why Goldman thinks that if it is worse to kill someone than to allow him to die this can *only* be on account of the principle she quotes. It is true that there is this principle, applicable only to examples in which harm is in question, and depending (I believe) on the idea that there is a more stringent duty to refrain from harming someone than to bring him aid. But there is another principle which has to do with rights, and which says that the right to non-interference extends further than the right to have one's purposes furthered by others. And since it is often possible to infringe someone's rights without harming him, and even in doing him good, this principle can and does affect the acceptability of acts of euthanasia.

The following example shows the principle at work in another area. A sees B drinking some harmful stuff and judges that he, A, has no right to take it from him; nevertheless when the bottle starts to roll away out of B's reach he refuses to put out his hand and get it back for him. Surely he will be right to refuse, for while B may be able to get another bottle, possibly he will not. To take the bottle from him would have been an impermissible paternalistic interference whereas it was a good thing to do to refuse to get it back. Other things are, or can easily be made to be, equal, and yet, since an impermissible act is worse than a good one, actively getting the bottle away is worse than allowing it to get away.

The application to euthanasia is obvious. Involuntary active euthanasia is never permissible because no one has the right to interfere with another's will in taking his life. (Or indeed to take it without his permission, actual or perhaps assumed.) It does not, however, follow that he must *serve* this will e.g. by continuing medical treatment in all circumstances. He has no right to interfere but he may have every right to refuse treatment which he sees as an evil to the patient. For a fuller discussion of this point see my 'Euthanasia', *Philosophy and Public Affairs*, winter 1977.

University of California at Los Angeles and
Somerville College, Oxford.

© PHILIPPA FOOT 1981

THEORIA

A Swedish journal of philosophy
Edited by Peter Gärdenfors, Bengt Hansson
and Krister Segerberg

Volume 45. part 3

Melvin C. Fitting An axiomatic approach to computers
Raimo Tuomela Putnam's realisms
Richard J. Hall Seeing perfectly dark things and the causal conditions of seeing
Agneta Sutton The Kantian and the consequentialist elements in Rawls's theory of justice

THEORIA, which invites contributions in all fields of analytic philosophy, usually publishes three issues a year. Subscriptions and requests for back numbers should be sent to *Administrative Editor,* THEORIA, *Filosofiska institutionen, Kungshuset i Lundagård, S-223 50 Lund, Sweden.* Subscription for one year is SwCr 55:– or U.S. $ 15.–, for three years SwCr 150:– or U.S. $ 38.–.

INQUIRY

Editor: Alastair Hannay

Vol. 23, No. 4, December 1980

Dangerous Knowledge? The Self-subversion of Social Deviance Theory	Terence Ball
On Robert Nozick's 'On Austrian Methodology'	Walter Block
Schemata in Social Science. Part One: Structural and Operational	J. O. Wisdom
Defending Marx's Theory of History (G. A. Cohen, *Karl Marx's Theory of History: A Defence)*	Leon Pompa
Ulysses and the Sirens (Jon Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality)*	F. H. Hahn

Vol. 24, No. 1, March 1981

Schemata in Social Science. Part Two: Metatheoretical	J. O. Wisdom
Hegel's Notion of *Aufheben*	B. C. Birchall
On Justifying Violence	Kai Nielsen
Reply to Kai Nielsen	H. J. N. Horsburgh
Intentions and Conditions of Satisfaction	Arthur R. Miller
Nagel on Subjective and Objective	Vinit Haksar
Apel and the Transcendental Pragmatics of Human Action (Karl-Otto Apel, *Die Erklären: Verstehen-Kontroverse in transzendental-pragmatischer Sicht)*	John M. Connolly

Subscription price:
Institutions: N.kr 140,– $28.00 – Individuals N.kr 84,– $17.00.

Published quarterly by Universitetsforlaget, P.O. Box 2959 Tøyen, Oslo 6, Norway – P.O. Box 258, Irvington-on-Hudson, New York 10533, USA.

11 JAN 1982

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

PUBLISHER. ANALYSIS is published by Basil Blackwell Publisher Ltd., 108 Cowley Road, Oxford OX4 1JF, England. US Mailing Agent, Expediters of the Printed Word Ltd., 527 Madison Avenue, Suite 1217, New York, NY 10022, USA. Second Class postage paid at New York.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £8.00 (inland), £9.80 (overseas), $24.00 (U.S.A.), $28.00 (Canada); for individuals £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A.), $17.50 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to the publisher.

CONTRIBUTIONS. **Articles submitted for publication should be addressed to Mr. Christopher Kirwan, Exeter College, Oxford OX1 3DP.**

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

PRINTED IN GREAT BRITAIN BY BURGESS & SON (ABINGDON) LTD., ABINGDON, OXFORDSHIRE

ol. 41 No. 4 (New Series No. 191) October 1981

ANALYSIS

Edited by
CHRISTOPHER KIRWAN

CONTENTS

The Wiener-Kuratowski procedure and the analysis of order — HERBERT HOCHBERG
Plantinga on the essence of essence — CAROLINE J. SIMON
Wittgenstein's operator *N* — P. T. GEACH
Was there ever a radical mistranslation? — IAN HACKING
Reply to Yu — JAMES HUDSON *and* MICHAEL TYE
On Yu on Hudson & Tye — WAYNE WASSERMAN
Expressions and tokens — PHILIP HUGLY *and* CHARLES SAYWARD
'Entailment' survives Lewy's paradoxes — JOHN K. SLANEY
Direct tableaux proofs — B. H. SLATER
Brand on event identity — PETER M. SIMONS
Russell on the reduction of particulars — ALBERT CASULLO
Hess on reasons and causes — PETER SMITH
Conditional murderers — CHRISTOPHER BRYANT
Butler's problem again — ERIC STIFFLER
Steiner on Nozick on the right to enforce — PATRICK WILSON
Steiner on rights and powers — DAVID MILLER
A negation-free version of the Berry paradox — ROBERT E. KIRK

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

BASIL BLACKWELL

Logical Positivism
OSWALD HANFLING

This book includes discussion of most of the main ideas of Logical Positivism, a prominent twentieth-century form of Empiricism advocated by the philosophers and scientists of the 'Vienna Circle' in the 1920s and '30s. Connections are made throughout with the classical empiricism of the seventeenth and eighteenth centuries, and with the ideas of Ludwig Wittgenstein.
190 pages, hardback £14.00 (0 631 10861 0) paperback £5.50 (0 631 12853 0)

Essential Readings in Logical Positivism
Edited by OSWALD HANFLING

A collection of readings to accompany *Logical Positivism*, this book includes some of the key writings of the Logical Positivists. The main writers represented are Moritz Schlick, Rudolf Carnap and Friedrich Waismann, and there are also items from G. E. Moore, A. J. Ayer, David Rynin, Hans Reichenbach and Carl Hempel.
256 pages, paperback £5.95 (0 631 12566 3)

Wittgenstein's Tractatus: An Introduction
H. O. MOUNCE

The aim of this book is to help the student with only an elementary knowledge of logic and philosophy to understand the main doctrines of Wittgenstein's *Tractatus*. It follows the order in which Wittgenstein wrote, and keeps technicalities to a minimum. H. O.Mounce covers everything of essential importance in the *Tractatus*, and also much of the more incidental detail.
144 pages, hardback £12.00 (0 631 12556 6) paperback £4.50 (0 631 12707 0)

Ludwig Wittgenstein: Personal Recollections
Edited by RUSH RHEES

Personal memories of the most important philosopher of this century by five people who each knew him intimately and in different ways: his sister Hermine Wittgenstein; his Russian teacher Fania Pascal; F. R. Leavis, a fellow Cambridge don; John King and M. O'C.Drury, two of his pupils.
248 pages, £9.50 (0 631 19600 8)

Wittgenstein and his Times
Edited by BRIAN McGUINNESS

The theme of this book is the agreement and difference between Wittgenstein's thought and that of those who were, in a broad sense his contemporaries. It contains five original contributions: Anthony Kenny on Wittgenstein and the Nature of Philosophy; Brian McGuinness on Freud and Wittgenstein; J. C. Nyiri on Wittgenstein's Later Work in relation to Conservatism; Rush Rhees on Wittgenstein on Language and Ritual; and G.H. von Wright on Wittgenstein in Relation to His Times.
128 pages, £7.95 (0 631 11161 1)

Basil Blackwell Publisher, 108 Cowley Road, Oxford OX4 1JF

BASIL BLACKWELL

New in Paperback

Philosophy and the Mirror of Nature
RICHARD RORTY

'This is an ambitious and important book. Ambitious because it attempts to place the main concerns and discussions of contemporary philosophy within a historical perspective; important because this is all too rarely attempted within our present philosophical culture, and almost never done this well . . . immensely rich and interesting book.' Charles Taylor, *Times Literary Supplement*
418 pages, paperback £4.95 (0 631 12838 7)

Naming and Necessity
SAUL A. KRIPKE

'. . . brilliant and very influential . . . stands up as an impressive and enduring work of philosophy,'outstanding in its sweep, clarity and penetration.' Colin McGinn, *Times Higher Education Supplement*
(July) 180 pages, paperback £3.50 (0 631 12801 8)

Sameness and Substance
DAVID WIGGINS

'There is no denying the importance of having a proper theory of identity. The attempt which David Wiggins has made to supply this need and the thorough honesty of this treatment . . . deserve to be treated with very great respect.' A. J. Ayer, *London Review of Books*
250 pages, paperback about £4.95 (0 631 12846 8)

Virtues and Vices
PHILIPPA FOOT

'. . . an important work. It is definitely one of the best available collections of modern ethical essays.' *New Blackfriars*
240 pages, paperback £4.50 (0 631 12749 6)

Posthumous Writings
GOTTLOB FREGE

This book contains all of Frege's extant unpublished writings on philosophy and logic, other than the correspondence. It comprises twenty-seven pieces, dating from 1879 to 1924-25.
(July) 302 pages, paperback £4.95 (0 631 12835 2)

Basil Blackwell Publisher, 108 Cowley Road, Oxford OX4 1JF

MIND A QUARTERLY REVIEW OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XC No 360 October 1981

Articles

- **The Question of Well-being and the Principle of Equality:** T. HONDERICH
- **Art and Sociobiology:** P. ZIFF
- **Literary Aesthetics and Literary Practice:** S. H. OLSEN
- **Mathematical Knowledge and Reliable Authority:** C. A. J. COADY
- **Harman's Paradox:** T. SORELL

Discussions

- **To Save Verisimilitude:** J AGASSI
- **Prescriptivism and Incompleteness:** S. B. TORRANCE
- **Yet More on Non-epistemic Seeing:** E. WRIGHT
- **On the Interpretation of Wittgenstein:** G. McFEE
- **Inconsistency and Contradiction:** J. N. WILLIAMS

Critical Notice

- **D. Wiggins, *Sameness and Substance*:** P. F. STRAWSON

Book Reviews

Books Received

Orders to:
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

Annual subscription:
Institutions: £10.00 or $22.50 (Canada $26.00)
Individuals: £8.00 or $18.00 (Canada $21.00)
Single issues: £3.00 or $6.75 (Canada $7.75)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

Philosophy

Published for the Royal Institute of Philosophy
Edited by Renford Bambrough
President of St. John's College, Cambridge

The journal is concerned with the study of philosophy in all its branches: logic, metaphysics, epistemology, ethics, aesthetics, social and political philosophy and the philosophies of religion, science, history, language, mind and education.

The journal is not committed to any particular school or method and contributors are expected to avoid needless technicality. There is a section on new books which includes reviews, book notes and a list of books received.

Volume 56 (1981) £23.00 per year £7.00 per part

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU, England
32 East 57th Street, New York, N.Y. 10022, U.S.A.

ANALYSIS 41:4 OCTOBER 1981

THE WIENER-KURATOWSKI PROCEDURE AND THE ANALYSIS OF ORDER

By HERBERT HOCHBERG

RUSSELL once dismissed the Wiener-Kuratowski procedure for the construal of ordered pairs as classes (viz. $<x, y> =$ def. $\{\{x\}, \{x, y\}\}$) as a trick. It is not clear what he meant,but I will argue here that there is a sense in which he was right: the use of the procedure does not eliminate the appeal to order.

The procedure enables one to introduce the pattern '$<x, y>$' so that (C) '$<x, y> = <w, z>$ iff ($x = w$ & $y = z$)' holds. Thus, it suffices for a treatment of ordered pairs in set-theoretical contexts. But there is an implicit claim in such a treatment—that the procedure suffices for an *analysis* of order. For, an ordered entity (an ordered pair) has purportedly been construed as an unordered entity (a class).[1] One is tempted to say that it obviously has since the *signs* '$\{\{a\}, \{a, b\}\}$' and '$\{\{a, b\}, \{a\}\}$' are signs for the same class, while '$<a, b>$' and '$<b, a>$' are signs for different ordered pairs. Kuratowski obviously did two things. First, he showed how to replace a sign that makes use of the linear ordering of constituent signs with a sign (a set sign giving a class in extension) that does not.[2] Second, he showed how to take an ordered pair as a class so as to satisfy (C). Classes are considered to be unordered even though their members may be given an order. Consider the class $\{a, b\}$. To say that it is not ordered is to say that $\{a, b\} = \{b, a\}$. To say that the members may be given an order is to say that some procedure may be used to establish an order among them; as trivial a procedure as enumerating them in an order. Recall Russell's well-known contrast of the class of all pairs of shoes with the class of all pairs of socks. *Being a left-shoe* could be used as a selection property. Thus, for such a class, if infinite, a "selection" axiom need not be appealed to in holding that there is a class consisting of one element from each member of the class of all pairs of shoes. Consider, next, the class of all classes of two members, such that one member is a pair-class while the other is a unit class.

---

[1] To give an analysis, as I take that notion here, is to show how one (supposed) kind of thing is to be taken in terms of another kind: a number as a class, a physical object as a class of sense data, a property as a class of particulars, a mental state as a physical state, a person as a collection of mental acts, and so on. In all such attempts it is understood that one cannot appeal to the kind being analysed in setting forth the analysans. Thus, one cannot take a person to be the class of all mental acts *had by the person.*

[2] The sign '$\{a, b\}$' involves a linear order, but in that '$\{a, b\}$' and '$\{b, a\}$' are signs for the same class we may say the linear ordering is not made use of.

Being a unit class is a property that can function, as Russell's *being a left-shoe* functions, to order the elements of each such pair-class. Moreover, in a perfectly clear sense it is unlike the property of *being a left-shoe* in that it may be called a logical or set-theoretical property. Where the members of a class may be ordered by employing such a property, I will speak of the class *being logically ordered.* The Wiener-Kuratowski procedure may be looked at as appealing to properties that serve to order classes logically. Is a class like $\{\{a\}, \{a, b\}\}$, then, an ordered entity? In that it is a class $\{\{a\}, \{a, b\}\} = \{\{a, b\}, \{a\}\}$, while, where $a \neq b$, $<a, b> \neq <b, a>$. Hence, we say that the class is not ordered in the sense in which $<a, b>$ is. But, the notion of being *logically ordered*, introduced above, points to an interesting way in which a class like $\{\{a\}, \{a, b\}\}$ differs from one like $\{a, b\}$. In the case of $\{a, b\}$ there is no set-theoretical or logical property that may be appealed to to order the elements of the set. This difference may seem insignificant, for one may suggest that in no sense can the set $\{\{a\}, \{a, b\}\}$ be taken as ordered since we can arbitrarily order the elements of the set in two ways by the use of the logical property *being a unit set.* Thus all one can do, as in the case of $\{a, b\}$, is impose an order. But, this points to the way in which sets like $\{\{a\}, \{a, b\}\}$ are *used* as ordered entities. Given the signs '$<a, b>$' and '$<b, a>$', one arbitrarily takes them to stand for one ordered pair rather than the other; i.e. one recognizes the ordering of the signs to correspond to the ordering of the elements. Likewise, one arbitrarily construes the ordered pair $<a, b>$ in terms of a set like $\{\{a\}, \{a, b\}\}$, or a variety of other alternatives. Finally, if, following Kuratowski, one uses sets like $\{\{a\}, \{a, b\}\}$ and $\{\{b\}, \{a, b\}\}$, one *chooses* which of $\{\{a\}, \{a, b\}\}$ or $\{\{b\}, \{a, b\}\}$ is to represent $<a, b>$ and which will represent $<b, a>$. This shows that in the construal of $<a, b>$ as $\{\{a\}, \{a, b\}\}$ one implicitly takes a and b in an ordering, for one takes the element in the unit set as the *first* element. We can see that in a simple way. Consider '$\{x_1, x_2, \ldots, x_n\}$' to be a function sign standing for a function, f_1, that yields a class as value for arguments $x_1, x_2, \ldots, x_n$—the class whose members are the arguments. One may say that the function f_1 is not an *ordering function* in two senses. (1) A class, whose members are the arguments, is the value; hence any sign of the form '$\{x_\alpha, x_\beta, \ldots, x_\nu\}$', which is the result of a permutation of the signs for the elements of the class (the arguments) is such that $\{x_\alpha, x_\beta, \ldots, x_\nu\} = \{x_1, x_2, \ldots, x_n\}$. (2) Taking the arguments in any order to be operated on by the function yields the same value. Consider, next, a two-term function, f_2, represented by '$\{\{x_1\}, \{x_1, x_2\}\}$'. Take a and b as arguments. f_2 also yields a class as value. The function is not an ordering function and the value is not ordered in the sense that a class is the value, and, hence, $\{\{a\}, \{a, b\}\} = \{\{a, b\}, \{a\}\}$. But f_2 is an ordering function and the class may be considered ordered in that the same class does not result as the value when the arguments are taken in a different order. The class, $\{\{a\}, \{a, b\}\}$, is the value of an ordering function for arguments a and b, though it is not the value of an ordering function for arguments $\{a\}$ and $\{a, b\}$, in the sense in which I am using the notion of an *ordering function*.

It will not do to suggest that we consider $\{\{a\}, \{a, b\}\}$ as the "result" of two applications of the class function f_1 so that the first application yields the classes $\{a\}$ and $\{a, b\}$ and another yields the class $\{\{a\}, \{a, b\}\}$. For, the first application does not employ the function f_1, since, given a and b as arguments, we must take them in an order to get $\{a\}$ and $\{a, b\}$, rather than $\{b\}$ and $\{a, b\}$. Thus, if we consider $<x_1, x_2>$ to be a function that yields an ordered pair for arguments a and b, such a function is like f_2, and not like f_1, in that it is an ordering function in the above sense. The point can be made more emphatic if we take f_2 to be a three-term function, which yields $\{\{a\}, \{a, b\}\}$ when the sequence (a, a, b) is taken as argument, and $\{\{b\}, \{a, b\}\}$, with (b, a, b) as argument. We may then note that we have sequences that differ not only in the arrangement of terms but in the repetition of *different* terms. These *two relevant differences* point to the implicit appeal to order in the *use* of a class like $\{\{a\}, \{a, b\}\}$.

It will also not do to argue that the point suggested here is vacuous, in that no function which will not be an ordering function in the above sense could be used to yield a set that would suffice for the construal of an ordered pair. The objection merely serves to emphasize a clear sense in which order has been presupposed and not analysed by Kuratowski's procedure. If the objection is pushed to suggest that on my argument no analysis of relational order could be presented, I think the point made is correct, but neither a *reductio* nor an objection.

To object that the Wiener-Kuratowski procedure has nothing to do with a purported *analysis* of order, but merely with *modelling* relational statements, avoids the problem. That one can provide a model of a certain kind is not at issue; just as it is not generally at issue that Russell provided a model of the Dedekind-Peano postulates. To provide a model is not to furnish an analysis.[3] Likewise, that alternative models exist does not imply that one, or more than one, model cannot be used as a basis for an analysis. The point here is not that there are alternative models of the Wiener-Kuratowski type, but that in constructing a model along such lines one implicitly appeals to an ordering of the elements a and b in the use of sets like $\{\{a\}, \{a, b\}\}$.

University of Texas

© HERBERT HOCHBERG 1981

---

[3] On this point see my 'Peano, Russell, and Logicism,' ANALYSIS, 16.5, April 1956, pp. 118–20.

PLANTINGA ON THE ESSENCE OF ESSENCE

By Caroline J. Simon

HAECCEITISM might be defined, very broadly, as the doctrine which holds that individuals can exist in more than one possible world, that the same individual can exhibit extreme dissimilarities from world to world, and that distinct individuals can be very similar. An Haecceitist is one who holds that this doctrine is true. This doctrine rests on the idea that each individual has a property or properties which he, she or it has necessarily and everything else necessarily lacks. This property is each individual's essence or haecceity. Haecceitists hold that, because each individual has an haecceity, there is a truth to the matter of whether a particular individual exists in a particular world and that truth in these matters cannot be determined simply by degrees of similarity.

Plantinga fits the above definition of an Haecceitist. He, for example, thinks that 'Socrates and Xenophon could have been such that the latter should have resembled Socrates as he was in the actual world more than the former' is true (Alvin Plantinga, *The Nature of Necessity*, Clarendon Press, 1974, p. 110; hereafter referred to as *NN*). Plantinga also uses haecceities and world-indexed properties to solve problems which arise concerning transworld identity. One of the implications that can be drawn from a recent paper by Robert Adams, 'Primitive Thisness and Primitive Identity,' (*Journal of Philosophy*, January 1979, pp. 5–26; hereafter referred to as *Adams*), is that Plantinga cannot both hold that haecceities are analysable in purely qualitative terms and employ world indexed properties in defending transworld identity.

The claim that Plantinga cannot both hold that haecceities are non-primitive and employ them in the way he does is interesting in light of the fact that he expresses an explicitly agnostic attitude toward the primitiveness of haecceities.

> Is each essence equivalent to some intersection of "qualitative" or "natural" properties? Is it possible that you should have all my qualitative properties and I all yours? Are there worlds that differ solely by a permutation of individuals? These questions are equivalent and equivalently difficult; I do not know their answers, partly because of uncertainties about the alleged notion of a qualitative property. (*NN*, p. 100)

The aim of this paper is to argue that if individuality can be understood in purely qualitative terms, Plantinga cannot successfully appeal to haecceities and world-indexed properties to solve problems raised by Chisholm concerning transworld identity.

Roderick Chisholm has raised a serious problem for transworld identity in 'Identity Through Possible Worlds: Some Questions' (*Nous* 1967, pp. 1–8; hereafter referred to as *Chisholm*). This problem arises out of the transitivity of the identity relation. Chisholm's example assumes that Adam and Noah are two actual, distinct individuals. If the idea of identity through possible worlds makes sense, there is surely a possible world W_1 which is just like α (where α is the actual world) except that Adam in W_1 has one insignificant property that Noah has in α and Noah in W_1 has one insignificant property that Adam has in α. This can be repeated through indefinitely many possible worlds until Adam and Noah have exchanged names, histories, physical characteristics, etc. The possible world thus reached, call it W_n, is just like α except "Adam" in W_n can be traced back to Noah in α and "Noah" in W_n can, be traced back to Adam in α. The exchange is so gradual that it is obvious in each case that Adam in α is identical to Adam in W_1, Adam in W_1 is identical to Adam in W_2, and so on. By the transitivity of the identity relation, "Adam" in W_n is identical to Noah in α and "Noah" in W_n is identical to Adam in α. However, it would appear that "Adam" in W_n is indiscernible from Adam in α and "Noah" in W_n is indiscernible from Noah in α. By the Identity of Indiscernibles, "Adam" in W_n is thus identical to Adam in α and "Noah" in W_n is identical to Noah in α. But, since identity is transitive, it follows that Adam in α is identical to Noah in α. Since this contradicts the original assumption that they are distinct, something has gone wrong.

Chisholm thinks that the root of the problem is the assumption that individuals exist in more than one possible world (*Chisholm*, p. 4). Those who wish to defend transworld identity need to make a case for the falsity of some other assumption made in the above case. There seem to be three alternatives: (1) one can maintain that "Adam" in W_n is discernible from Adam in α, because they have different world-indexed properties, (2) one can maintain that Adam in α is not identical to "Noah" in W_n, because somewhere in the transition one of the properties exchanged was essential to Adam, or (3) one can maintain that the Identity of Indiscernibles is false. However, if Plantinga were to hold that essences are analysable in purely qualitative terms, these alternatives would be either unavailable or ineffective.

Plantinga might find (1) very attractive, for he frequently appeals to world-indexed properties in the process of solving transworld identity problems (e.g. in 'Transworld Identity or Worldbound Individuals?' in *Naming, Necessity, and Natural Kinds,* ed. Stephen P. Swartz, Cornell University Press, 1977, p. 255). Take any property, *P*, that Adam has uniquely in α: *having P in* α will individuate Adam in every world in which he exists. For example, if *being the first human* is a property unique to Adam in α, *being the first human in* α should be a property which Adam in α has but "Adam" in W_n lacks. Likewise *being the builder of the ark in* α would make Noah in α discernible from "Noah" in W_n. However, given a purely qualitative view of individuals and possible worlds this will not work. Having a property in α comes to having a property in a world with suchness S_a (i.e. in a world with all the qualitative properties which the actual world has). Since, by hypothesis, the only difference between α and W_n is in which individuals instantiate which qualitative properties, α and W_n are qualitatively indiscernible. If this is the case, having a property in W_n *also* comes to having a property in a world with suchness S_a. Adam's property *being the first human in* α is thus the property of *being the first human in a world with suchness* S_a, and "Adam" in W_n also has this property. Therefore, if possible worlds are purely qualitative, world-indexed properties cannot be used to individuate Adam in α from "Adam" in W_n.

Plantinga might protest that this conclusion is far too hasty. Adam in α *is* discernible from "Adam" in W_n as can be seen by focusing on world-indexed properties which they instantiate in worlds other than α and W_n. Take, for example, one of the indefinitely many worlds in the transition between α and W_n, W_{979}, a world in which Adam uniquely instantiates *being the first human and being the builder of the ark*. *Being the first human and being the builder of the ark in* W_{979} is a property which Adam has in α, but "Adam" lacks in W_n. Moreover, since on the qualitative analysis of individuals and possible worlds, this is a purely qualitative property of Adam, α and W_n *are* qualitatively discernible. Part of α's suchness is that the person who has all of Adam's actual suchnesses (i.e. those qualitative properties he has in α, considered without their world-indexes) would have instantiated *being the first human and being the builder of the ark* if W_{979} had been actual. Part of W_n's suchness is that the person who would have instantiated all of Adam's actual suchnesses had W_n been actual, would not have instantiated *being the first human and building the ark* if W_{979} had been actual. Thus, Adam is qualitatively discernible from "Adam" and W_n is qualitatively discernible from α, if one takes into account world-indexed properties.

However, this reply is also inadequate if possible worlds are purely qualitative. This is because there is a possible world, W_{979}*, which is just like W_{979}, except that in W_{979}* it is "Adam" in W_n who instantiates *being the first human and being the builder of the ark*. In general, for every world-indexed property, *P in W*, which one might try to claim that Adam in α has but "Adam" in W_n lacks, "Adam" has an analogue of the property, *P in W**, where the only difference between W and W* is that, in W, Adam has *P* and, in W*, "Adam" has *P*. If this is true, the suchnesses of W and W* are indiscernible. So, *being the first human and being the builder of the ark in* W_{979} comes to *being the first human and being the builder of the ark in a world with suchness* S_{979}. Since having the analogue of this property in W_{979}* comes to the same thing, this is not a property that Adam in α has but "Adam" in W_n lacks. So, to repeat our conclusion, it follows that if possible worlds are purely qualitative, world-indexed properties cannot be used to individuate Adam in α from "Adam" in W_n.

Chisholm suggests that (2) might be a possible, though not unproblematic option (*Chisholm*, p. 6). We might want to say that at some point one of the properties we took from Adam was an essential property of his. In other words, although the changes can be as slight as one wishes in the transition from α to W_n, at some point Adam exists in W_{m-1} but does not exist in W_m. Even if this were a possible alternative for Plantinga, other things he has said indicate that he would not be likely to adopt it. For one thing, Plantinga expresses doubt that any "ordinary meat-and-potatoes" properties (or any disjunction of them) are essential to an individual (*NN*, p. 61). What Plantinga seems to mean is that there are no non-world-indexed properties which are essential to a particular human being, for example, but which are not essential to all human beings. So, there should be no properties of Adam's which Noah cannot instantiate and vice versa. Moreover, for it to be the case that Adam exists in W_{m-1} but not in W_m, there would have to be a reason why it is impossible for him to have the suchness he would instantiate if W_m were actual in a world with W_m's suchness, given his essential properties. What kind of a reason could there be for this?

The only alternative left is (3). One can argue that even though "Adam" in W_n is indiscernible from Adam in α, they are not identical, because the Identity of Indiscernibles is false. But, as Robert Adams has persuasively argued, if one holds that essences are suchnesses, it follows that the Identity of Indiscernibles is true. If what it is to be a particular individual is just to have a certain collection of qualities (world-indexed or not), then the only sense that can be made of being a numerically different individual is having different qualities. As Adams puts this point, 'If individuals are *infimae species*, then "the principle of individuation is always some specific difference."' (*Adams*, p. 12) Thus, Plantinga can take alternative (3) only by holding that essences are primitive.

It seems then that Plantinga either has to hold that essences are not suchnesses or concede that Chisholm's Adam-Noah case functions as a *reductio* which shows that transworld identity is impossible. Furthermore, Plantinga's use of world-indexed properties seems to rest on assuming essences are primitive. Thus, the question of whether essences are primitive is not one about which Plantinga can afford to be in doubt.

University of Washington, Seattle © CAROLINE J. SIMON 1981

WITTGENSTEIN'S OPERATOR *N*

By P. T. Geach

THE thesis that there is a class of propositions, the elementary propositions, from which *all* propositions can be derived by repeated application of the operator N (*Tractatus logico-philosophicus* 6.001) is manifestly difficult in various ways. I shall not here be discussing these difficulties. Before discussion that is worth anything can even start, people must know what the propositional operator N is and how it works: in much discussion of the *Tractatus* even this modest preliminary requirement is not fulfilled.

The N operator yields joint denial of an *arbitrary* number of propositional arguments, from 1 upwards. Applied to one argument it yields the same result as ordinary negation; applied to two arguments it yields binary joint denial; but it may be applied to more than two arguments, and then these need not be listed, but may be given by specifying the class containing just such-and-such propositions. Wittgenstein spells all this out quite unequivocally (5.501); all the same, his meaning has often been mistaken.

One mistake is simply to confuse the N operator with binary joint denial. Historically, I conjecture, this error arose from a careless reading of Russell's introduction to the *Tractatus* (p. 13f.). The reading must have been *very* careless, since Russell there explicitly speaks of 'simultaneous denial of all the propositions which are values of fx'. Nevertheless, the mistake is common. In a recently published mimeographed volume of papers on Wittgenstein, there was frequent reference to the N operator under the style 'the Sheffer stroke'; nobody at the conference, it appears, pointed out the error. And in Favrholt's volume on the *Tractatus* it is taken for granted that the N operator cannot be applied to more than two propositions: if e.g. there are four propositions then we must select two out of the four before applying the operator; 'Wittgenstein has apparently not thought about this peculiarity'! (Favrholt, p. 134).

The degree of logical incompetence Favrholt ascribes to Wittgenstein is even greater than this: he speaks of Wittgenstein's 'mistake' of not seeing that when we negate p, q, we do not arrive at $\sim p . \sim q$ but at $\sim p$, $\sim q$ . . . 'To conjugate (*sic*) these two results, as Wittgenstein does, is not permissible'. (p. 32) Favrholt goes on to say (*loc. cit.*): '. . . if the negation of p, q is $N(p, q)$, then the negation of $N(p, q)$ must be p, q . . . My conclusion is that Wittgenstein has made two serious mistakes. First, he has not observed that p, q is not a truth-function and that therefore it cannot be negated as a whole. Secondly, he has overlooked the fact that the expression $N(N(p, q))$ is equivalent to p, q and not to

$p \vee q$.' Of course the 'serious mistakes' here are Favrholt's own; he fails to see that though the N operator takes a class of propositions as its basis, it results not in a class of propositions but a single proposition: thus '$N(N(p,q))$' is the 'joint' denial of a *unit* class of propositions; but anyhow a joint denial must be a *single* proposition and cannot, as Favrholt seems to think, be a pair of propositions.

To bring out in full the way Wittgenstein's N operator works, we need (something he does not himself provide) an explicit notation for a class of propositions in which one constituent varies. I shall write '$N(\ddot{x} : fx)$' to mean joint denial of the class of propositions got by substituting actual names for the variable in the propositional function (represented by) 'fx'. Thus '$(\exists x)fx$' and '$(x)fx$' will come out in a Wittgenstein-style notation as '$N(N(\ddot{x} : fx))$' and '$N(\ddot{x} : N(fx))$' respectively. The first is the denial of the joint denial of the class of propositions saying that such-and-such is f, and thus says that something or other is f: the second is the joint denial of the class of propositions saying that such-and-such is *not* f, and this amounts to saying that nothing is not f, i.e. that everything is f.

Clearly, once we have such a notation for a single quantifier, we can repeat the procedure to symbolize a proposition which in the familiar Russellian notation has any number of prenex quantifiers. '$(\exists x)(y)(fxy)$' will come out, for example, as follows:

$$N(N(\ddot{x} : (N(\ddot{y} : N(fxy)))))$$

This looks complicated, but we can easily break it down. '$N(\ddot{y} : N(fay))$' says that we are to deny every such proposition as '$N(fab)$', i.e. that 'fab' is always to be affirmed whatever 'b' stands for; i.e., it is equivalent to '$(y)fay$'. And then if we replace 'a' by the variable 'x' and embed the result in '$N(N(\ddot{x} : (\ldots)))$', we shall be saying that the joint denial of the class of propositions we get from '$(y)(fay)$' by varying 'a' is itself to be denied, i.e. that one or other member of this class is to be affirmed, i.e. that '$(\exists x)(y)(fxy)$' is to be affirmed.

Notice that although the class-of-propositions notation is iterated, we do not get classes *of classes* of propositions, but only classes of propositions. Wittgenstein was exaggerating when he said that the theory of classes is altogether superflous in mathematics (6.031) for he cannot get on without classes of propositions. But his set theory is very rudimentary: it has no null class, nor has it classes of classes; and by the same token it can never generate any Russellian paradoxes. About the set theory in the *Tractatus*, the excuse that it is only a very little one may reasonably be heard from a defender of Wittgenstein.

A recent book on Wittgenstein by Robert Fogelin claims that the *Tractatus* theory of quantification breaks down already when we have as many as two prenex quantifications to deal with: Fogelin calls this (it is the title of his chapter VI, §1) 'a fundamental error in the logic of the *Tractatus*'. Fogelin might have reflected that Russell and Ramsey were unlikely not to test the *Tractatus* theory for the multiply quantified cases, unlikely not to detect a breakdown, and unlikely to have taken the *Tractatus* as seriously as they did if there were a breakdown. Of course the 'fundamental error' is Fogelin's, and is not hard to detect. Fogelin thinks there is a difficulty about e.g. negating each proposition in an infinite set and then jointly negating the results: this would involve, before the final joint negation, 'completing a task involving infinitely many discrete steps'. *Whose* task, one might ask, is this one? *Nobody* has to run through an infinite series of propositions, negating them one by one, and *then* perform a further task of joint denial; if A specifies the truth-conditions of a proposition by reference to an infinite class of propositions, each of which is the negation of some proposition, this specification does not impose upon A the task of *himself* constructing all the negations in 'infinitely many discrete steps'. If Fogelin were right, Wittgenstein's theory of quantification would break down in any infinite Universe even for *singly* quantified propositions, e.g. for '$(x)fx$', which comes out in my suggested notation as '$N(\ddot{x} : N(fx))$'.

I end by illustrating how the N operator would work in harness with Wittgenstein's notation for a formal series. The notation

$$[a\mathrm{R}b,\ a\mathrm{S}b,\ a\mathrm{R}/\mathrm{S}b]$$

gives us the series of propositions

$$a\mathrm{R}b,\ a\mathrm{R}/\mathrm{R}b,\ a\mathrm{R}/(\mathrm{R}/\mathrm{R})b,\ \textit{etc. ad inf.}$$

Using the N operator then, we get a proposition

$$N[a\mathrm{R}b,\ a\mathrm{S}b,\ a\mathrm{R}/\mathrm{S}b]$$

to say that *a* is *not* an R-ancestor of *b* at any remove; accordingly

$$N(N[a\mathrm{R}b,\ a\mathrm{S}b,\ a\mathrm{R}/\mathrm{S}b])$$

says that *a does* bear *some* power of the relation R to *b*. This is expressed in Russell and Whitehead's notation by '$a\mathrm{R}_{\mathrm{po}}b$', not by '$a\mathrm{R}_{*}b$'; the latter involves counting identity (in a restricted field) as the initial generation of ancestorhood, but that is ruled out by the *Tractatus* theory (stated in the 5.53 sections) that identity is *not* a relation.

University of Leeds © P. T. GEACH 1981

REFERENCES

[1] D. Favrholt. *An interpretation and critique of Wittgenstein's TRACTATUS* Copenhagen, Munksgaard, 1967.

[2] R. Fogelin: *Wittgenstein* (in *The Arguments of the Philosophers series*, ed. Ted Honderich, Routledge and Kegan Paul, 1976).

Addendum. Since writing this article I find that the confusion between the *N* operator and the Sheffer stroke occurs also in H. O. Mounce's book *Wittgenstein's Tractatus, An Introduction* (Blackwell 1981), pp. 52–3: admittedly there is some modification of the misstatement later on in the chapter on generality.

WAS THERE EVER A RADICAL MISTRANSLATION?

By IAN HACKING

> On their voyage of discovery to Australia a group of Captain Cook's sailors captured a young kangaroo and brought the strange creature back on board their ship. No one knew what it was, so some men were sent ashore to ask the natives. When the sailors returned they told their mates, 'It's a kangaroo.' Many years later it was discovered that when the aborigines said 'kangaroo' they were not in fact naming the animal, but replying to their questioners, 'What did you say?'[1]

THAT would be a radical mistranslation if the story were true. A radical mistranslation is defined as follows: (1) Speakers of two very different languages are trying to communicate. (2) A speaker of one language says *s*. Speakers of the other language take him to be saying *p*. (3) This translation is completely wrong. Yet (4) neither party realizes it, although they continue to converse. Moreover (5) the mistranslation persists until it is too late to correct.

Condition (1) restricts us to people who are talking to each other and excludes the decoding of ancient texts. Condition (3) rules out mere differences of nuance, moderate misunderstandings and misclassifications that occur to all of us all the time. Since my story is about naming I shall call this kind of radical mistranslation *malostension.* That occurs when (6) an expression of the first language is taken by speakers of the second language to name a natural kind. (7) It does nothing of the sort, but (8) the second language incorporates this expression as the name of the natural kind in question. (7) is intended as a strong condition. Malostensions are not just misclassifications, or the taking of the name of an individual as the name of a class.

I cannot prove that radical mistranslations never occur. But I shall show that some famous alleged malostensions are frauds, founded upon rumour and refuted by facts. This may matter, because of the doctrine of the indeterminacy of translation. Quine's thesis is a priori, but it gains credence partly from anecdotes. We tend to feel that indeterminacy is radical mistranslation carried to the limit, where no possible information can settle which of two incompatible translations is correct. I reject this plausibility argument by rebutting the anecdotes with which it begins. The remarkable thing about human interaction is that mistranslations are so readily cleared up.

---

[1] Quoted from *The Observer* (London) 1973 on p. 150 of my *Why Does Language Matter to Philosophy?* (Cambridge University Press, 1975). Many thanks to Professor J. J. C. Smart for telling me that the story is false, and for providing me with references 2 and 3.

Cook's team recorded many words of a language they encountered in the Endeavour River area of Australia. They were confident of the spelling and meaning of only 60 of these words. 'Kangaroo' was among them. Later travellers did not in fact encounter the word. Hence the story with which I began—people made up an explanation of Cook's word. But the story is based only on the fact that few subsequent travellers spoke with the Australian community that Cook had met. This was apparently pointed out in a letter to an Australian newspaper in 1898, but did not become common knowledge until the work of the anthropologist John Havilland in 1972.[2] He wrote up a vocabulary for a dialect called Guugu Yamidhirr, spoken by people in just the area that Cook landed. Their word for kangaroos is "ga*n*urru," where I use "*n*" for a phoneme which is a bit like "ng." Kangaroo, in short, was no malostension, although perhaps our spelling is the result of poor phonetics.

There is a printed conversation in which Quine and Putnam mention two other alleged malostensions:

> Quine: David Lewis pointed out a nice example to me [. . .] There was, in the nineteenth century, a French naturalist named Pierre Sonnerat, who was doing field work in Madagascar. A lemur went up a tree, and Sonnerat asked a native "Qu' est que c'est?" The native said "in dri," which in Malagasy means "There he goes." Sonnerat thought that the native understood his question and had given the answer, and the animal is known as the indri to this day.
> Putnam: That's like the word 'vasistas'.
> Quine: Right. The French word for transom is "Was ist das".[3]

Putnam and Quine do not actually assert that *vasistas* is a malostension, but others have told me that it is. Yet it is hard to imagine the circumstances. Were some German tourists pointing at a French transom, asking 'Was ist das?' Did some French people overhear that and think, 'Hah, that little window must be called a *vasistas*'? One can hardly credit that. In fact the French word was first written, in French, with German spelling. It is thus used by the French mining engineer J. F. C. Morand in a 1776 book chiefly describing foreign equipment for mines. He speaks of a *Wass ist das* and means a little grille, built into a door and which can be opened to see what is outside or pass small objects, messages, etc., in or out.

---

[2] R. M. W. Dixon, *The Languages of Australia* (Cambridge University Press, 1980) pp. 8–9.

[3] *Synthese* 27 (1974) p. 500.

Thus the 'What is that?' does not have the force of a question, namely, what is this funny little window or opening. Rather *vasistas* is a word intended to convey a function. The *vasistas* is the grille through which you look to find out 'What is that?' i.e., what is that thing outside, who is knocking, who is there?

In 1784 one finds *Wasistdas* as the name of a proposed small window in an enclosed platform attached to a Mongolfier balloon. The passengers could be completely protected, but could open the *Wasistas* from time to time to see the view.[4] The spelling of the word soon became completely gallicized. The *vasistas* had been a grille to separate the house from outsiders but by 1793 popular justice gave it the reverse role. The lunette in the guillotine was jocularly called the vasistas as in: *Passer la tête au vasistas*! No malostension here.

Indri is a little trickier. I suspect that we have exactly the same situation as with 'kangaroo,' namely a word used in a dialect encountered by Sonnerat, but not noticed by later travellers. Beware of thinking that such writers were careless. Cook was reporting back to Sir Joseph Banks, one of the most critical Newtonian inductivists of all time. Cook did not rely on casual reports of his sailors (as in the story with which I began). His team sifted all their data and settled on only 60 words, including 'kangaroo' of which they were sure. Likewise Sonnerat was no tripper watching lemurs scoot up trees (as Quine's version tends to imply). He was one of the most detailed of reporters and it is on work like his that Cuvier was glad to rely.

The Oxford English Dictionary does assert that 'indri' is a malostension. The evidence is provided by a missionary from Madagascar writing, in 1893, in a magazine that he edited. Rev. Sibree says of the short-tailed lemur that,

> their native name is *Bàbakòto,* literally 'Father-child' (or boy), not *Indri,* as said by Sonnerat, who discovered the species. *Indri* (or *indry*) is a Malagasy word meaning 'lo!' or 'behold!' and was probably mistaken by him for a name and other Europeans for a name, when the natives exclaimed: "Indry izy!" ("There he is!")[5]

Was this then the perfect malostension? First let us get our authorities straight. All modern etymological dictionaries fall back on the old Oxford English Dictionary and its citation of Sibree. But in fact the above quotation from Sibree is an unacknowledged word-for-word translation of the observation made in 1868 by François Pollen, another missionary and naturalist.[6] Now whether we owe the story to Rev. Sibree or Father Pollen, Quine has certainly captured its spirit. We imagine that Pierre Sonnerat visits Madagascar in 1780 and asks what a certain lemur is called. He gets *indri* or somesuch in reply, and the animal scoots up a tree. We get the picture of Sonnerat never getting much closer to an indri—they are forever getting away.

---

[4] Gunnar von Proschwitz, 'En marge du Bloch-Wartburg,' *Studia Neophilologica* 36 (1964) p. 329.

[5] J. Sibree, *Antanarivo Annual and Madagascar Magazine* 5 (1893) p. 83.

But in fact, as Sonnerat tells us[7], indri are very easily tamed, and in the Southern part of the island are used in hunting, much as we use dogs. Far from just overhearing a cry of indri he asks what it means, and is told that it means little man of the woods. So Sonnerat not only heard the word, but discussed it. He later printed his engraving of a nice fat indri eating a . . . banana? Might that be a fanciful reconstruction of a glimpsed animal that he never got close to? No, he took one aboard ship and later presented it to his king. For a while it played in the royal garden and was later stuffed and put in the Paris natural history museum. You may see it there to this very day.

None of this proves conclusively that Sonnerat got the name right. You will not find the word *indri* in dictionaries translating from Malagasy into European tongues. But taking a hint from the kangaroo story, we may ask if Sonnerat got the phonemes slightly wrong. Sure enough: the nineteenth century dictionaries do have a word *endrina* for a species of lemur.[8] This might, of course, be a back-formation from French, but given that Sonnerat was visiting a very big island that had long been on Arab and then on European trade routes, and given that he was a gifted naturalist with able companions, it seems likely that we can rely on his report. Even if the word *endrina* was not widespread, the people he talked to used something like *endrina* for the lemur in question. Once again, there was no malostension.

Certainly there are errors of understanding that do persist, but they contrast with what I call radical mistranslation. Here are two examples.

---

[6] F. P. L. Pollen, 'Notices sur quelques autres mammifères habitent Madagascar et les îles voisines,' in *Recherches sur la Faune de Madagascar et de ses dépendances d'après les découvertes de François P. L. Pollen et D. C. van Dam*. 2me partie. *Mammifères et oiseaux par H. Schlegel et François P. L. Pollen.* Leiden 1868, p. 20.

[7] Pierre Sonnerat, *Voyage aux Indes Orientales et à la Chine, fait par ordre du Roi depuis* 1774 *jusqu'en* 1781, Paris, 1782, Vol. II pp. 141–143, and Plate 88. Second edition, 1806, Vol. IV, pp. 89–92. *Collection de planches pour servir au voyage aux Indes Orientales et à la Chine,* Paris, 1806, Planche 86.

[8] The first dictionary for Malagasy and a European language is *A Dictionary of the Malagasy Language,* Part I, English-Malagasy, by J. J. Freeman; Part II, Malagasy sy English, by D. Johns; Antanarivo, London Missionary Society, 1835. Freeman provides no entries in English for lemurs. Johns does have the word *endrina* which he translates as a kind of monkey. The next English dictionary is *A New Malagasey English Dictionary*, J. Richardson, Antanarivo, London Missionary Society, 1885. Here *endrina* is translated as a kind of lemur and other sources confirm this. We infer that the earlier English missionaries did not see that the lemur is not a kind of monkey, but, at the time they were writing, *endrina* was current in Malagasy. (At that time there were virtually no true monkeys on the island.) Sibree's own copy of the Freeman dictionary is now in the library of Yale University. It has alternating blank pages so that the user can enlarge the list of entries in the field. Sibree made few additions, and none of a naturalistic sort.

Among the Malagasy words that now apply to lemurs we find *gidro* and *rajako* (or *jako*). A gidro is a smallish grey lemur. The word is derived from the Swahili *ngedere* which means monkey. There are no lemurs in most of Africa, nor were there any monkeys native to Madagascar. So it was some sort of mistake when traders speaking Swahili called a lemur by the name of *ngedere*, but it was an error of classification, not of translation. As for *jako*, it too, ought strictly to apply only to monkeys. An English sea captain had a pet monkey that escaped and the sailors ran after it, calling it by name. 'Jack! Jack!' they cried, and the name stuck.[9]

Stanford University

© IAN HACKING 1981

---

[9] That, at any rate, is what Richardson's 1885 dictionary says.

REPLY TO YU

By James Hudson *and* Michael Tye

IN 'The Modal Argument Against Description Theories of Names' (ANALYSIS 40.4, October 1980) Paul Yu rejects an argument against the theory that ordinary proper names are covert definite descriptions, an argument which we sketched rather informally in 'Proper Names and Definite Descriptions with Widest Scope' (ANALYSIS 40.1, January 1980). Perhaps the informality of our presentation led to Yu's misunderstanding of the argument; the present note should clarify the matter beyond question.

Our first point is that the whole issue of whether names can be viewed as definite descriptions is an issue about the *meanings* of names, about their contribution to the tasks performed by the sentences in which they occur. We assume, then, the elementary distinction between a symbol and its meaning, or between a sentence and the proposition which it (normally) expresses. In what follows we shall use this distinction without apology, since however talk of propositions may be *analysed*, it surely cannot be altogether illegitimate.

Now let us assume that Russell's Theory of Descriptions is correct in its application to explicit definite descriptions. The description theory of names holds that ordinary proper names can be viewed as covert definite descriptions; that for each name n there is a (possibly complex) predicate F such that n is equivalent to $\ulcorner(\iota x)Fx\urcorner$ as analysed by Russell.

Kripke sought to refute this theory by considering the behaviour of names in modal contexts. In particular he claimed that $\ulcorner\Diamond Fn\urcorner$, where F is the predicate of the description supposed equivalent to n, is not in general equivalent to $\ulcorner\Diamond(\exists x)(y)(Fy \equiv x = y)\urcorner$, as it should be on the description theory. But the description theory thus refuted equates a name with a description accorded less than the maximum scope, so it would still seem possible that each proper name be equivalent to a definite description *which is always to be accorded the widest possible scope in any sentence in which it occurs.*

To refute this last and most plausible version of the descriptionist theory we examined (*op. cit.*) the behaviour of names in non-modal sentences. We claimed that pairs of non-modal sentences $\ulcorner Fn\urcorner$ and $\ulcorner F(\iota x)Fx\urcorner$ (the latter equivalent to $\ulcorner(\exists x)(y)(Fy \equiv x = y)\urcorner$) are not synonymous; that the proposition expressed by the former is not entailed by the proposition expressed by the latter (actually the converse entailment does not hold, either). Scope considerations are irrelevant here, since there is no scope ambiguity about $\ulcorner F(\iota x)Fx\urcorner$.

Yu tries to refute our claim of non-entailment by showing it to be equivalent to another claim which the descriptionist could plausibly reject, for reasons having to do with scope ambiguity. Our claim (labelled 2 (ii) in Yu's paper) was essentially this:

> (1) The proposition expressed by $\ulcorner(\exists x)(y)(Fy \equiv x = y)\urcorner$ does not entail the proposition expressed by $\ulcorner Fn\urcorner$.

Yu equates this with:

> (2) The proposition expressed by $\ulcorner\Box[(\exists x)(y)(Fy \equiv x = y) \supset Fn]\urcorner$ is not true

(Yu's 3 (ii)). Ordinarily one would accept such an equation, on the principle that:

> (3) The proposition expressed by S_1 entails the proposition expressed by S_2

amounts to:

> (4) The proposition expressed by $\ulcorner S_1 \supset S_2\urcorner$ is necessary,

and thus to:

> (5) The proposition expressed by $\ulcorner\Box(S_1 \supset S_2)\urcorner$ is true.

But the reason the move from (3) to (5) seems attractive is that one tacitly assumes that each of the sentences S_1 and S_2 is in all contexts the self-contained expression of a single proposition. Allow names, construed as widest-scope descriptions, to occur in either sentence and this assumption will prove false. Instead the scope of a name occurring within a sentence will be sensitive to the wider context in which the sentence is placed, so that, e.g., the scope of a name within S_1 standing alone will differ from its scope in $\ulcorner S_1 \supset S_2\urcorner$, and both will differ from its scope in $\ulcorner\Box\ (S_1 \supset S_2)\urcorner$. Thus the move from (3) to (4) will be unacceptable, as one can see by letting $S_1 =$ 'Homer was blind' and $S_2 =$ 'Someone was blind', where 'Homer' is assumed equivalent to 'the author of the *Iliad* and the *Odyssey*'. (Then (3) is the true claim that there being a unique author of the *Iliad* and *Odyssey*, who was blind, entails that someone was blind; and (4) is the false attribution of necessity to the proposition that there is a unique author of the *Iliad* and the *Odyssey* who is such that, if he was blind, someone was blind.) And the move from (4) to (5) is no better, for again the scope of the alleged definite description (disguised as a name) will differ from one context to another. It is ironical that the central argument of Yu's paper in favour of the widest-scope description theory uses reasoning which he ought to reject precisely because he holds that very theory.

Our guess is that Yu went wrong because he was not careful to distinguish between sentences and propositions (a distinction which, for expository convenience, we soft-pedalled in our original note). In any case he says nothing to undermine our refutation of the theory that names are definite descriptions with widest possible scope.

Northern Illinois University © JAMES HUDSON *and* MICHAEL TYE 1981

ON YU ON HUDSON & TYE

By WAYNE WASSERMAN

IN ANALYSIS 40.1 (January 1980, pp. 63–64) Hudson and Tye argued against the view which treats names as wide scope occurrences of definite descriptions. Paul Yu criticized their argument in ANALYSIS 40.4 (October 1980, pp. 208–209). I intend to explain why I am less than fully satisfied with that criticism.

Hudson and Tye began by reviewing the classical modal argument against the descriptivism which identifies names with definite descriptions. The argument is familiar, and I shall simply refer to it as (1). (I shall refer to arguments with the same numbers as Yu). They then observe that descriptivists have avoided the force of (1) by modifying their position to the view that names are wide scope occurrences of definite descriptions. Hudson and Tye in turn modify (1) in an attempt to refute the new version of descriptivism. Their argument runs like this. The descriptivism in question must treat 'N is F' and 'the F is F' as synonymous, for 'the F' clearly has wide scope in the latter. Since synonymous sentences have the same entailment relations, and since 'there is a unique F' entails 'the F is F', the descriptivism must grant that 'N is F' is also so entailed. This provides the basis for a modus tollens:

(2) (i) If *N* is synonymous with wide scope occurrences of *the F*, then 'N is F' is entailed by 'there is a unique F'.
(ii) 'N is F' is not so entailed.
(iii) Thus *N* is not synonymous with wide scope occurrences of *the F*.

Yu responds to this argument by considering another argument which he urges is equivalent to (2). The argument is formed by substituting 'necessarily, if there is a unique F, then N is F' for the consequent of premise (i), and by making a similar change in premise (ii). Since the substituting phrase contains a modal operator, and 'since names are usually read with widest scope,' Yu employs the descriptivist's identification of names with wide scope occurrences of descriptions to claim that the consequent of (i) is equivalent to both of the following: 'N is such that, necessarily, if there is a unique F, then N is F'; 'the F is such that, necessarily, if there is a unique F, then the F is F'. So read (i) is false and (2) is unsound.

Suppose it is granted that there is a reading of the consequent of (i) which is false, and thus that there is a reading on which 'there is a unique F' does not entail 'N is F'. Since synonymous sentences have the same entailment relations, and since 'N is F' and 'the F is F' are by hypothesis synonymous, it would follow that there is a reading of 'the F is F' which is not entailed by 'there is a unique F'. Nonetheless, there seems to be a reading of 'the F is F' which is so entailed. There would then seem to be two senses of 'the F is F'. The first, which for obvious reasons I shall call the rigid reading, is not entailed by 'there is a unique F'. The second, which I shall call the non-rigid reading, is so entailed. (There would thus be two senses of ' "there is a unique F" entails "the F is F" ', one true and one false.) This poses a dilemma for the descriptivism which identifies names with wide scope occurrences of definite descriptions. Either 'N is F' is synonymous with only one of those readings, the rigid one, or it is also ambiguous between a rigid and a non-rigid reading. In the former case the consequent of (i) has only the false reading considered by Yu, in the latter case it is ambiguous between a true and a false reading.

Consider first the suggestion that 'N is F' is synonymous only with the rigid reading of 'the F is F'. Then the descriptivism at issue has been abandoned; N is no longer synonymous with any wide scope occurrence of the F, but only with rigid wide scope occurrences. Besides requiring an additional and controversial ambiguity between rigid and non-rigid senses of wide scope descriptions, the resulting view poses an additional problem. Scope considerations seem irrelevant. Why not simply identify names with rigidly designating descriptions, and, since a rigid description need not be one whose quantifier has widest scope, base replies to kripkean objections solely on considerations of rigidity and non-rigidity?[1] On the other hand, if it is held that both 'N is F' and 'the F is

---

[1] Leonard Linsky (*Names and Descriptions*, University of Chicago Press, 1977) distinguishes between considerations of scope and rigidity and non-rigidity of designation, and notes on page 59 that rigid descriptions need not have widest scope. In 'Speaker's Reference and Semantic Reference' (reprinted in *Contemporary Perspectives in the Philosophy of Language*, edited by Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, University of Minnesota Press, 1979) Kripke discusses the ambiguities which allegedly infest sentences containing definite descriptions, and he too carefully distinguishes between scope considerations and considerations of rigidity and non-rigidity.

F' are ambiguous between rigid and non-rigid readings, then the original motive for adopting the modified descriptivism vanishes. The unmodified descriptivism was abandoned because names and descriptions allegedly behaved differently; names were supposed to be rigid and descriptions were not. This alleged difference is rejected if both 'N is F' and 'the F is F' are held to be ambiguous between rigid and non-rigid readings. Therefore, in this case too it is unclear why scope is relevant to descriptivism. Once it is granted that the interesting ambiguity is that between rigid and non-rigid designation, and that there is no obvious connection between that and ambiguity of scope, scope considerations appear irrelevant.

An advocate of the view that names are synonymous with wide scope occurrences of definite descriptions can deny that there is any sense of 'the F is F' which is entailed by 'there is a unique F', and thereby avoid the dilemma. However, given the currency of the view that there is such a sense, this denial requires defence. It is hard enough to argue that there is a rigid as well as a non-rigid sense of definite descriptions in such sentences as 'the F is G' and 'the F is F'. To claim that there is only a rigid sense requires considerable support. In addition, if there is no non-rigid reading of 'the F is F', then both the original argument advanced by Kripke and Hudson and Tye's modification rest on a shared misunderstanding of definite descriptions. There would then be no reason to think that Kripke's argument requires the rejection of the original descriptivism, and the proper response to both Kripke and Hudson and Tye is to reject both arguments on the grounds that they presuppose a faulty account of definite descriptions.

Despite Yu's useful comments, I conclude that Hudson and Tye have advanced considerations sufficient for rejecting the view that names are wide scope occurrences of definite descriptions.[2]

University of Nebraska © WAYNE WASSERMAN 1981

---

[2] I thank Philip Hugly, Albert Casullo, and the Editor and a referee of ANALYSIS.

EXPRESSIONS AND TOKENS

By PHILIP HUGLY *and* CHARLES SAYWARD

THIS paper performs the limited but useful task of uncovering and correcting several confusions about expressions, tokens and the relations between them that crop up in even highly sophisticated writing about language and logic.

I

Joseph Schoenfield ([4], p. 2) says that 'an axiom (or theorem) may be viewed in two ways. It may be viewed as a sentence, i.e., as the object which appears on paper when we write down the axiom, or as the meaning of a sentence . . . '. He chooses to deal with axioms and theorems as sentences, and one reason he gives is that 'a sentence . . . is a concrete object; so that by studying axioms as sentences, we approach the abstract through the concrete'.

Schoenfield confuses sentences with sentence-tokens. On the one hand, he is saying a sentence is a concrete object, a visible series of marks on paper. On the other hand, he regards the set of sentences of his formal language *L* ([4], section 2.4) to be provably infinite.

Sentences, and expressions generally, are abstract objects; for no other construal will permit a proof that a language has infinitely many sentences. Tokens (e.g. sentence tokens) are concrete objects—perceptible particulars. One can write down sentences. But the result of writing down a sentence is, e.g., a visible spread of ink, a sentence token. The result of writing down a sentence is not the sentence itself. The sentence itself is imperceptible.

That expressions are abstract entities while tokens are concrete entities is generally recognized by people who think about such matters. But even people who grasp the basic distinction get confused about important details.

Susan Haack is an example. She writes:

> A sentence token is a physical object, a series of marks on paper or of sound waves . . . One could think of a sentence type either as a pattern which similar tokens exemplify, or as a class of similar tokens. ([1], p. 75)

Consider the two characterizations of sentence types Haack proposes:

(H1) A sentence type is a pattern which similar tokens exemplify.
(H2) A sentence type is a class of similar sentence tokens.

We are not sure what Haack means by a pattern; perhaps, she means a two or three dimensional geometrical shape. If so, a sentence is an abstract entity. Fine. The trouble comes with her saying the shape is exemplified by similar tokens. If the sentence type is untokened, it is not exemplified by any token. The same difficulty comes out more clearly with (H2). From (H2) it follows that any two untokened sentence types ('sentences' for short) are identical.

Again, Benson Mates writes:

> A sentence, at least in its written form, is an object having a shape accessible to sensory perception, or, at worst, it is a set of such objects. ([2], p. 10)

This seems doubly wrong. First, it is the token that has the perceptually accessible shape, not the sentence it tokens. Second, the sentence, whatever it is, is not the set of its tokens, for the reason recently alluded to in connection with Haack.

For a final example of confusing expressions with tokens we turn to Quine's device of quasi-quotation. Quine introduces this device to avoid certain errors inherent in expressing metalinguistic generalizations through quotation. Here is a standard instance:

(a) If ϕ is a sentence, so is '$\sim\phi$'

which is, by the standard account of quotation names,[1] a vacuous quantification of the false sentence

(b) '$\sim\phi$' is a sentence.

Thus, (a) also is false. Quine recommends replacing (a) by

(c) If ϕ is a sentence, so is $\ulcorner\sim\phi\urcorner$,

of which an instance would be,

(d) '$\sim$Fa' is a sentence

for '$\sim$Fa' as value of 'ϕ'. We immediately note that though (a) is an error, it typically leads to no further errors. Generations of logicians knew perfectly well to infer from (a) that '$\sim$Fa' is a sentence if 'Fa' is one and to *not* infer (b) from (a). Similarly, we know how in practice to work with the corners. The question before us is whether Quine gives a coherent and error free account of the function of the corners.

---

[1] Note our use of quotation marks is standard. Quotation names name expressions, not their tokens. E.g. 'Fa' = 'Fa'.

Quine introduces the concept relative to

(3) $\ulcorner(\mu)\urcorner$, $\ulcorner\phi \equiv \psi\urcorner$

in the following way:

> The quasi-quotations (3) are synonymous with the following verbal descriptions:
>
> The result of writing '(' and then μ and then ')',
> The result of writing ϕ and then '$\equiv$' and then ψ;
>
> or, equivalently:
>
> The result of putting μ in the blank of '( )',
> The result of putting ϕ and ψ in the respective blanks of ' $\equiv$ ';
>
> or, equivalently:
>
> The result of putting μ for 'μ' in '(μ)',
> The result of putting ϕ for 'ϕ' and ψ for 'ψ' in '$\phi \equiv \psi$'.
>
> . . . Described in another way: a quasi-quotation designates that (unspecified) expression which is obtained from the contents of the corners by replacing the Greek letters . . . by the (unspecified) expressions which they designate. ([3], p. 36)

According to the first sentence of this explanation (c) is to be read as

(e) If ϕ is a sentence so is the result of writing '$\sim$' and then ϕ.

This confuses types with tokens. The result of writing something is a concrete entity. A sentence is an abstract entity.

Essentially the same point can be made about each of the other formulations which Quine holds to be equivalent to the first. For note that *putting* and *replacing* are as much activities resulting in tokens as writing. When a person replaces 'The' by 'A' in 'The house is red' he, e.g., erases the token of 'The' and produces in its place a token of 'A'; or, perhaps, he simply writes 'A house is red'.

In one sense the above difficulty is easily eliminated. Simply read the corners as:

The expression tokened by the result of writing . . .

But note that since there are but finitely many tokened negations of a first order language *L*, (c) fails to yield *L*'s infinitely many negations under the above reading.

A solution of sorts to this difficulty is also at hand, if we read the corners as follows:

> The expression that would result were one to write . . . [2]

But Quine would reject this reading due to its modal character. And it does seem plain that first order languages should not require for their specification an intensional metalanguage.

In our view quasi-quotation has no place in metalinguistic descriptions of first order languages. The underlying difficulty is the attempt to use a tokening terminology, which makes reference to finitely many concrete acts productive of finitely many concrete objects, to specify infinite structures. The attempt is just another manifestation of the failure to sharply distinguish expressions, which are abstract and infinite in number, from their tokens, which are concrete and finite in number.

II

Questions remain. It is all very well to be told that an expression is an abstract entity, but what kind of an abstract entity is it? Is an expression a class? If so, what are its members? Is it a geometrical shape? Two dimensional? Three dimensional? How are syntactical descriptions of even very simple languages to be given so as to be in accord with the general idea that an expression is an abstract entity?

We think expressions are well construed as classes. The primitive expressions of the language can be construed as classes of perceptible particulars similar in some physical respect to given perceptible particulars. The rest of the expressions of the language are classes built up out of these classes. We hope to make this basic view plausible by showing how a syntactical description of a first order language can be given in accord with it.

In what follows 'PS' abbreviates 'is physically similar in shape to'. Consider these ink marks:

(2) x
(4) P
(5) ′
(6) N
(7) ∃

Now consider these sets:

A = {x : x PS the ink mark at (2)}
B = {x : x PS the ink mark at (4)}

---

[2] We owe this suggestion to the Editor.

$C = \{x : x \text{ PS the ink mark at } (5)\}$
$D = \{x : x \text{ PS the ink mark at } (6)\}$
$E = \{x : x \text{ PS the ink mark at } (7)\}$

The primitive signs of the language (call it 'X') are the five sets A, B, C, D and E.

Variables. (i) $<A, C>$ is a variable. (ii) if α is a variable so is $<\alpha, C>$. (iii) Nothing else is a variable.

Predicates. (i) B is an O-ary predicate. (ii) If Π is an n-*ary* predicate, $<\Pi, C>$ is an n+1-*ary* predicate. (iii) If Π is an m-*ary* predicate, so is $<C, \Pi>$. (iv) Nothing else is an m-*ary* predicate.

Formulas. (i) If $\alpha_1, \ldots, \alpha_m$ are variables and Π is an m-ary predicate, $<\Pi, \alpha_1, \ldots, \alpha_m>$ is a formula. (ii) If $<a_1, \ldots, a_h>$ and $<b_1, \ldots, b_i>$ are formulas, $<D, a_1, \ldots, a_h, b_1, \ldots, b_i>$ is a formula. (iii) If α is a variable and $<a_1, \ldots, a_h>$ is a formula, $<E, \alpha, a_1, \ldots, a_h>$ is a formula. (iv) Nothing else is a formula.

Sentences. For any formula ϕ and positive integer i, let ϕ_i be the i-*th* element of ϕ. For any formula ϕ, variable α, positive integers i, m and n: ϕ_i is a bound occurrence of α in ϕ iff ϕ_i is α and $<\phi_m, \phi_{m+1}, \ldots, \phi_{m+n}>$ is a formula and ϕ_m is E, ϕ_{m+1} is α and for some $h(1 \leqslant h \leqslant n)$ ϕ_i is ϕ_{m+h}; ϕ_i is a free occurrence of α in ϕ iff ϕ_i is α and ϕ_i is not a bound occurrence of α in ϕ. For any formula ϕ, variable α and positive integer i: ϕ is a sentence iff ϕ_i is not a free occurrence of α in ϕ.

The above definition of a bound occurrence of a variable may be usefully compared with the standard definition. Mates defines the concept in this way:

> .... an occurrence of a variable α in a formula ϕ is *bound* if it is within an occurrence in ϕ of a formula of the form $(\alpha)\psi$ or of the form $(\exists\alpha)\psi$; otherwise it is a free occurrence ([2], p. 45).

He further says: 'The variables are the lower case letters "*u*" through "*z*", with or without numerical subscripts . . . ' ([2], p. 44).
Now consider

(8) $(\exists\, u_1)\ F^1_1\, u_1$

While there *seem* to be two occurrences of '*u*' in (8) (as well as two occurrences of 'u_1'), neither occurrence of '*u*' is bound according to the definition. Mates says 'a sentence is a formula in which no variable occurs free' ([2], p. 45). So (8) is not a sentence. This is not the result he wants, and it raises havoc with the completeness proof (to take one example) if allowed to stand.

The difficulty is traceable to the notion of an occurrence. On a straightforward interpretation '*u*' occurs in (8) *since* 'u_1' does. According to the set-theoretic construction we have given, to say a variable α occurs in a formula ϕ is just to say, for some positive integer i, α is ϕ_i. So <A, C> might not occur in a formula even if <<A, C>, C> does.

III

What is to be said about *being a token of* which relates a concrete object (the token) to an abstract object (the expression)?

First, it does not just relate these two items; one has to at least add the language as a third item. But this isn't enough. Note that it is only because of the Morse Code that objects made up of dots and dashes are tokens of expressions of English. A perceptible particular is a token of an expression of a language L only relative to some tokening system. *Being a token of* is a 4-place relation between perceptible particular, expression, language and tokening system.

A tokening system is a set of instructions that tell how, for any given expression, a speaker of the language could construct a perceptible particular that tokens that expression given that the speaker had enough physical and mental resources (time, patience, chalk, memory capacity, space, and so on). This is a necessary condition; it is not sufficient. For consider the set of instructions whose only element is this:

> For any expression of English, token that expression by inscribing a mark similar in shape to this one:
>
> Z

This does tell a speaker how to token expressions of English. But it doesn't tell a hearer which particular expression the speaker has tokened. For a speaker S to communicate with an audience A by using a language, S has to produce perceptible particulars *and* A has to recognize these particulars as tokens and determine which unique expression each tokens. So, in addition to revealing how a speaker can construct tokens, a tokening system must also reveal how a hearer can process tokens. Putting both points together we can say: A tokening system must reveal how one can *unambiguously* token particular expressions.

Consider the language X described in the previous section. We are sure that, given any expression of X, we could unambiguously token that expression given sufficient mental and physical resources. We may not have the memory capacity or patience or space, etc., to carry this out in a particular case; but we know how to go about it. A construction of a tokening system for X is a formal representation of this capacity.

In the case of X the construction might go as follows:

Tokening the primitive signs. To token A construct a mark physically similar in shape to the mark at (2); similarly for B—E.

Tokening variables. To token $<\alpha, C>$, where α is a variable, first token α and then follow it a small but discernible distance to the right by the result of tokening C.

Tokening predicates. (i) To token $<C, \Pi>$ where Π is a O-ary predicate first token C and then follow it a small but discernible distance to the right by the result of tokening Π; clauses (ii) and (iii) are similar.

Tokening formulas. (i) To token $<\Pi, \alpha_1, \ldots, \alpha_m>$, where Π is an m-ary predicate and $\alpha_1, \ldots, \alpha_m$ are variables, first token Π and then follow it a small but discernible distance to the right by the result of tokening α_1, and then, . . . , and then follow α_{m-1} a small but discernible distance to the right by the result of tokening α_m; clauses (ii) and (iii) are similar.

A tokening capacity is formally represented by a set of recursively specified sets of expressions together with a set of tokening rules couched in the terms of those recursions. Given a recursive specification of a set of sets of expressions (= a language abstractly construed), it is easy to provide adequate tokening rules, as in our example. But it is not always easy to provide an adequate set of tokening rules e.g., for a natural language such as English. The difficulty, however, lies not so much on the side of the tokening rules themselves as on the side of the recursive specifications of the sets of expressions constituting the natural language abstractly construed.

The University of Nebraska—Lincoln

© PHILIP HUGLY *and*
CHARLES SAYWARD 1981

REFERENCES

[1] Susan Haack, *Philosophy of Logics*, Cambridge University Press, 1978.
[2] Benson Mates, *Elementary Logic*, Second Edition, Oxford University Press, 1972.
[3] W. V. O. Quine, *Mathematical Logic*, Revised Edition, Harper and Row, 1951.
[4] Joseph Schoenfield, *Mathematical Logic*, Addison-Wesley, 1967.

'ENTAILMENT' SURVIVES LEWY'S PARADOXES

By JOHN K. SLANEY

CASIMIR LEWY puts forward in [2] arguments purporting to show that various proposed analyses of logical entailment and of conditionals fall short of adequacy. One of his main targets is the theory due to A. R. Anderson and N. D. Belnap and their followers. There are two main thrusts to his attack. The first is an appeal to the supposed plausibility of the "disjunctive syllogism" in its truth-functional form—

$$(A \vee B) \; \& \; {\sim}A \vdash B$$

—and I have no wish here to add to the ink that has already been spilt over that topic. The second takes the form of four semiformal arguments which he claims to be intuitively invalid though sound on any appropriate extension of the Anderson-Belnap systems. The claim is that Anderson and Belnap have misidentified the "paradoxicality" of "paradoxes of implication" since their own logic allows as valid what are intuitively fallacies of relevance. I shall essay a defence of what Lewy calls 'the Anderson-Belnap school' against this second line of criticism.

Lewy's *Paradox* 1 concerns the sentences:

(A) The proposition expressed by the English sentence 'There is no brother who is not male' is the proposition that there is no brother who is not male.

(B) The proposition that there is no brother who is not male is necessary.

(C) The proposition expressed by the English sentence 'There is no brother who is not male' is necessary.

The derivation (inessentially shortened) goes:

(1) A & B entails C
(2) A is contingent
(3) B is necessary
(4) C is contingent
(5) therefore A entails C
(6) A entails A & C (from (5) and simple properties of '&')
(7) A & C entails B
(8) therefore A entails B.

But (8) is paradoxical in the sense of being counter-intuitive. Paradox 1 is to be found on pages 117–118 of [2].

Let us pretend there is nothing problematic about modal logic or propositional identity, and concentrate on the relevantly invalid step (5). The general principle of suppression of necessary premisses may be stated

A & B entails C
B is necessary
therefore A entails C.

Lewy appears to agree with Anderson and Belnap that this leads to intuitively unacceptable results, but he claims that the weaker version used at line (5) 'seems intuitively acceptable', as it appeals also to the contingency of A and C. But let P and Q be any propositions, Q necessary and P contingent. Then

P & Q entails P & Q
P is contingent
Q is necessary
P & Q is contingent
therefore (?) P entails P & Q
but P & Q entails Q
therefore P entails Q.

If the weakened suppression principle seems intuitively acceptable to Lewy then he must accept that 'It is raining' *intuitively* entails '$2 + 2 = 4$' on the grounds that the weather could be fine. Anderson and Belnap, of course, will have none of it. Lewy's claim (on page 131 of [2]) that the Anderson-Belnap theorists cannot give a general definition of entailment which will avoid line (8) of paradox 1 is utterly without foundation or support. Appeals to the necessity of B, with or without the contingency of A and C, are precisely what Anderson and Belnap reject as sources of entailment; for them paradox 1 is not a problem.

Paradoxes 2 *and* 3 can be dismissed very quickly. They use the same derivation as paradox 1 except that they replace A, B and C by

(A′) Caesar is dead $\equiv$ Russell is a brother;
(B′) Russell is a brother $\equiv$ Russell is a male sibling;
(C′) Caesar is dead $\equiv$ Russell is a male sibling;

and

(A″) Caesar is dead $\equiv$ This has shape;
(B″) This has shape $\equiv$ This has size;
(C″) Caesar is dead $\equiv$ This has size;

respectively. Clearly the derivations are invalid in suitable extensions of the Anderson-Belnap systems for the same reason as before. Moreover, if by '$\equiv$' Lewy means the truth functional

$$P \equiv Q = (P \;\&\; Q) \vee (\sim P \;\&\; \sim Q)$$

then even step (1) fails relevantly, since Anderson and Belnap inveigh against supposing that material equivalence is a kind of equivalence just as much as they do against reading material implication as a kind of implication. I also note that paradox 3 is unfortunate in that (B″) might well be false (for instance if "this" is the number 17)—but let us not descend into quibbles.

The most interesting part of Lewy's attack is *Paradox* 4. Consider:

(A) There are exactly ten brothers;
(B) There are exactly as many brothers as brothers.
(X) P entails P & P;
(Y) 'There are exactly ten Fs and exactly ten Gs' entails
'There are exactly as many Fs as Gs'.

Now A entails A & A by X; and A & A entails B by Y. Hence A entails B. But this conclusion is rather counter-intuitive. Lewy asserts, without any proof, that

> '. . . . Paradox 4 cannot be avoided on the lines of the Anderson-Belnap approach.' ([2], p. 131)

He also asserts that the argument does not involve illegitimate suppression of any kind. Let us see.

Consider

(C) There are exactly 10 Fs and exactly 10 Gs;
(D) There are exactly as many Fs as Gs.

For a theorist of the Anderson-Belnap school, D is *ambiguous*. It might be read as any of:

(D1) For some n there are exactly n Fs and exactly n Gs;
(D2) For every n there are n Fs if and only if there are n Gs;
(D3) For some m and n there are exactly m Fs, there are exactly n Gs, and m = n.

There may be further readings, depending on the resources of the theory in which all this takes place. And the paradoxicality or otherwise of paradox 4 depends on whether and in what sense the inference from C to D is valid.

Relevantly and intuitively, C does entail D1, for ten is as good an instance of n as any. Hence insofar as we read D as D1 there just is no paradox. But C does not relevantly (or, Anderson and Belnap would claim, intuitively) entail D2. Insofar as we read D as D2 there is a paradox, a paradox of implication, much criticised by Anderson and Belnap: that of deriving 'if P, Q' from 'P and Q'. Now D3 is a fair approximation to 'The number of Fs is (identical with) the number of Gs', which is a very natural thing to mean by D. It does not relevantly follow from C *alone*, for it requires the additional (suppressed) premiss '10 = 10'. That is, Anderson and Belnap, in line with their general policy, will allow the inference

There are 10 Fs & There are 10 Gs & 10 = 10
therefore
∃m∃n (There are m Fs & There are n Gs & m = n)

without being committed to the view that the last of the conjoined premisses is superfluous (even if '10 = 10' is a theorem of their theory of brother-enumeration).

Classically, of course, all of D1, D2 and D3 are simply equivalent; one of the main points of doing relevant logic is to be able to draw just this sort of classically impossible fine distinction. Even the above brief sketch of an analysis of paradox 4 shows, I think, how the Anderson-Belnap insights may be used to explain why we feel the inference from A to B is paradoxical on the one hand and acceptable on the other. Moreover, in the sense in which there *is* illegitimate suppression involved, we are shown what is suppressed and what forms of the conclusion its restoration would permit. The 'theorists of the Anderson-Belnap school' thus have a delicate, penetrating and intuitively satisfying analysis of paradox 4 while Lewy himself offers none. Paradox 4 is therefore a triumph and not a disaster for Anderson and Belnap.

In conclusion, I should say that I am not disposed to dispute Lewy's claim that our intuitions about logic are muddled, incomplete and downright inconsistent. Nor am I disposed to maintain that any of the Anderson-Belnap systems is "correct" and classical logic "incorrect" (or *vice versa*). I wish merely to have shown that Lewy's attack on 'the Anderson-Belnap school' is easily repulsed and that his four paradoxes will not support the weight of his argument.

University of Durham © JOHN K. SLANEY 1981

REFERENCES

[1] Anderson, A. R. & Belnap, N. D., *Entailment: the logic of relevance and necessity,* volume I, Princeton University Press, Princeton, 1975.

[2] Lewy, C., *Meaning and Modality,* Cambridge University Press, Cambridge, 1976.

DIRECT TABLEAUX PROOFS

By B. H. Slater

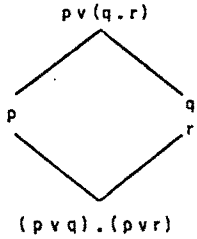
CLASSICALLY, valid arguments are just those whose premises are inconsistent with the denial of their conclusions. This is the basis, for instance, of the common method of establishing validity by means of semantic tableaux: one takes the premises and the negation of the conclusion and shows that all the branches in the resulting tableau close.

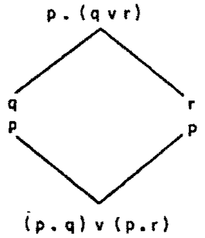
But there are doubts about the classical definition of validity. It leads, amongst other things, to the accreditation of 'p, -p ∴ q' (so contradictions entail anything) and much energy has been spent searching for an alternative definition which will avoid such results.

Given, however, the naturalness of the tableaux technique, it is surprising that no-one (to my knowledge) has tried this simple modification of it: one takes merely the premises of an argument and concludes what remains in those branches of the resulting tableaux *which do not close*. It is immediately obvious that, on this basis, contradictions will entail nothing. The rest of this short paper is devoted to describing some of the other main features of the entailment relation thus defined.

For ease of procedure I shall use the (propositional) tableaux in W. Hodges recent primer *Logic* (Penguin, 1977, p. 321). For ease of comparison I shall use (for reasons which will become apparent) the tautology-less system of deduction set out in I. Copi's *Symbolic Logic* (4th. ed. Macmillan 1973, Ch 3). It needs to be stated what conclusions may be read off the branches left open in a tableau—clearly, by reversing the rules for premises, we may derive from such branches any disjunction each of whose disjuncts is in some branch and for which each branch contains some disjunct (conjunctions of atomic sentences in a branch being counted as being in that branch) just as, in tabling a disjunction, we form branches such that each disjunct is in some branch and each branch contains some disjunct. Clearly, also, we may derive any conjunction of the disjunctions which are derivable.

It is fairly clear that all the usual Boolean equivalences Copi describes as 'Replacement Rules' (p. 37) are left standing. This follows from the way tableaux are normally constructed, for example with the laws of distribution we get the following:

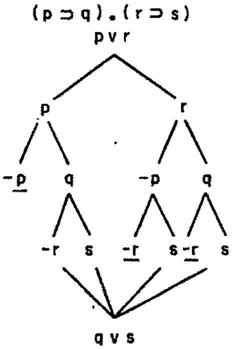




In each case the process is reversible (in the second case simply by going upwards, for 'p' is a degenerate disjunction whose single member spans the open branches).

Of course, Copi's 'Replacement Rules' are general inter-substitution rules, but there is no difficulty in seeing that, for instance, 'p v (q . r)' and '(p v q) . (p v r)' are to the same effect whenever they are in the same place in a formula.

Copi's 'Inference Rules' (p. 32) bring in the closure of certain branches of the tableaux. For instance, 'Constructive Dilemma' is proved as follows:



In this way we get *modus ponens*, *modus tollens*, Disjunctive Syllogism etc., and clearly Simplification and Conjunction—in fact all of Copi's 'Inference Rules' with one exception: Addition. There is no way with the present procedures to deduce 'p v q' from just 'p', since to read off 'p v q' as a conclusion, amongst other things, 'q' must be in some open branch, and this is not given.

The absence of Addition is essential to the avoidance of 'contradictions entail anything', since it is well known that Addition, with Simplification and Disjunctive Syllogism will give that result. Without Addition we clearly also avoid the standard 'paradoxes of material implication', and an easy treatment of counterfactuals becomes available: 'If it had been that p, it would have been that q' can now without difficulty be represented '-p . (p ⊃ q)' (I use the common symbol hook, but clearly it does not have all its usual significance).

The divergence of the present system of deduction from Copi's tautology-less one is therefore welcome yet minimal: to present the system in Copi's manner all we need do, in fact, is drop one rule. But its presentation in terms of tableaux would seem the most natural one, and all we need do then is change the usual definition of validity.

It remains to say a word about Conditional Proof, the addition of which gives Copi (pp 50–61) his tautologies, and turns all his inference rules into logical truths. It should be clear that we cannot add Conditional Proof to our present system without collapsing it into Copi's. This is for two reasons, at least. First, given Simplification we could get '$(-p \,.\, -q) \supset -p$' and hence '$p \supset (p \vee q)$' as tautologies, whence we could get Addition as a derived rule, by *modus ponens*. But also Conditional Proof is linked directly with Addition through such proofs as this:

1. $-p$	Assumption
2. q	Assumption
3. $-p \,.\, q$	From 1 and 2
4. q	From 3, and so from 1 and 2
5. $-p \supset q$	From 2 (using C.P. on line 4)
6. $p \vee q$	From 5, and hence from 2

Copi shows (p. 53) that Conditional Proof, with Addition (which is superfluous given the above) yields Indirect Proof: if from 'p' one can derive a contradiction, deduce '$-p$'. *That* is why Conditional Proof is so intimately linked with the standard system—it is what is behind the classical definition of validity. For if from '$P \,.\, -C$' one can derive a contradiction, that means, classically, we can get '$P \supset C$', and the tautologousness of that, with *modus ponens*, means '$P \therefore C$' is a valid argument. Hence it is entirely in tune with our rejection of the classical conception of validity that we should also reject Conditional Proof. (Of course if '$P \therefore C$' is valid still '$P \,.\, -C$' is contradictory, but the relation no longer holds in the reverse way).

What we are left with, therefore, is quite essentially a *logic without tautologies*, i.e. a logic consisting purely of inference and substitution rules (or their equivalent in tableaux terms) and no 'logical truths'. If there is anything to be said for calling Natural Deduction systems 'natural' it lies in this idea of logic as a body of inter-factual derivation rules rather than non-factual truths. But common systems of Natural Deduction include a Conditional Proof rule—and that, we can now see, distorts the whole picture.

University of Western Australia © B. H. SLATER 1981

BRAND ON EVENT IDENTITY

By Peter M. Simons

MYLES BRAND espouses a theory of events as dated, unrepeatable particulars, very much in the mould of Davidson. (M. Brand, 'Particulars, events, and actions', in M. Brand and D. Walton (eds.), *Action Theory*, Dordrecht: Reidel, 1976, pp. 133–157.) He disagrees with Davidson, however, as to the identity conditions of events. According to Davidson, events are the same which have the same causes and the same effects. (D. Davidson, 'The individuation of events', in N. Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel*, Dordrecht: Reidel, 1969, pp. 216–234, reprinted in D. Davidson, *Essays on Action and Events*, Oxford: Clarendon, 1980, pp. 163–180.) Brand criticises this on three grounds. Firstly, it is epistemically ineffective. That is to say, we must frequently know whether events are the same in advance of deciding whether or not they have the same causes and effects. This objection, while correct, is not telling, as it is epistemological rather than ontological. The second point is that the thesis depends on the contentious principle of universal causation. Were there two or more events lacking causes and effects, Davidson's criterion would falsely identify them. It would also falsely identify events occurring in causal chains which only temporarily diverge. The third objection is that it is illegitimate to have an identity criterion for events which itself quantifies over events. This objection was first made by Neil Wilson (N. Wilson, 'Facts, events and their identity conditions', *Philosophical Studies* 25 (1974), 303–321, esp. pp. 303–4) and seems the most effective of the three.

Brand accordingly undertakes to replace Davidson's criterion with one of his own, according to which events are the same which necessarily occur within the same spatio-temporal regions. The purpose of this note is simply to argue that Brand's criterion is itself unsatisfactory. I have no general criterion to put in its place: indeed I rather doubt if there is one.

Suppose, with Brand, that 'e' and 'f' are event-variables, 's' ranges over spatio-temporal regions, and R is the relation '. . . occurs within—'. The criterion

$$(1)\ e = f \text{ iff } (s)(eRs \equiv fRs)$$

is rejected by Brand, as by Davidson before him, because two different events may take up the same region, as when a metal sphere rotates while simultaneously getting warmer. So Brand strengthens the criterion to

$$(2)\ e = f \text{ iff } \Box(s)(eRs \equiv fRs)$$

The sphere's rotating is by this criterion different from its getting warmer, since either could have taken place without the other, so they do not necessarily coincide.

Necessary coincidence seems a good sufficient condition for event identity. But is it a necessary condition? I believe not. Brand is forced to modify (2) somewhat to take account of contingent identity sentences involving descriptions. For

(3) Nixon = the 37th President

is contingently true. Hence the sentence

(4) Nixon's resigning = the 37th President's resigning

is also true, and also contingently, just as 'Nixon's mother = the 37th President's mother' is contingently true. But Brand's criterion (2) makes (4) false, which it should not. To take account of this, Brand adopts the Kaplan *Dthat*-operator which turns descriptions into rigid designators of the objects actually designated by the descriptions. If α is any singular term, α^+ is the rigid term obtained from it by applying *Dthat* to all descriptions occurring in extensional positions within α (the restriction is essential if α and α^+ are to co-designate). Brand's revised criterion is then

$$\begin{aligned}(5)\ e = f &\text{ iff } e^+ = f^+\\ e^+ = f^+ &\text{ iff } \Box(s)(e^+ \mathrm{R} s \equiv f^+ \mathrm{R} s)\end{aligned}$$

and by this criterion (4) correctly turns out true.

Contingent identity (3) gave rise to the need to replace (2) by (5). But if (3) is contingently true, so is (4): note that it has the form of an identity containing two functional expressions, which are obviously equivalent to descriptions, so there is no question of (4) being necessary by virtue of the terms flanking the identity being rigid. Now, according to Brand, 'Identity conditions for events are *logically* necessary and sufficient conditions for the truth of sentences of the form '$e = f$'.' (Brand, *op cit.,* 144 (emphasis added).) This means that 'iff' in (5) must be interpreted in each case as strict equivalence. But then (5), together with the plausible S4 modal principle

$$(6)\ \Box p \supset \Box\Box p$$

yields, by distribution of necessity across strict equivalence, that $e = f \supset \Box(e = f)$, and hence (4) turns out to be necessarily true rather than contingently true as assumed. So something has to give: either the view that (4) is contingent, or the S4 principle (6), or Brand's criterion (5). Of the three, the last seems to me the weakest.

The first part of (5) is here the cause of trouble. Its right-hand side is by construction necessary, but its left-hand side should be contingent. Independently of questions of event-identity, the *Dthat*-operator turns contingent identities to necessary ones: 'Nixon = the 37th President' is contingently true, but 'Nixon$^+$ = (the 37th President)$^+$' is necessarily true. So the first 'iff' must be weakened to material equivalence. This means that (5) no longer offers identity conditions in Brand's sense of logically necessary and sufficient conditions, but at least it delivers up the right *truth-value* for (4).

So amended, the criterion takes account of the contingency of the identity (4), and appears to be an effective criterion. So, at least, it is treated by Brand, who uses it confidently to resolve questions of action-identity. For example, the identity

(7) Stravinsky's lowering the baton at t_0 = Stravinsky's signalling to the cellos at t_0

(where I have replaced Brand's '(now)' by 'at t_0') is declared by Brand to be false on the strength of (5). The connection between the actions, although they coincide, is taken by him to be one of convention rather than identity. This is a case where a supporter of Davidson would say that the actions are identical: Brand accepts that he here differentiates events which Davidson would identify.

The supporter of Davidson can however protest that the criterion (5), *as Brand uses it*, simply begs identity questions. The trouble concerns the application of the operator *Dthat* to singular terms. It was simple to decide, in respect of (4), that (Nixon's resigning)$^+$ is necessarily identical with (the 37th President's resigning)$^+$, because we know that (3) is true, and hence Nixon is necessarily identical with (the 37th President)$^+$. The identity (4) follows because the two sides of (4) both contain the same noun gerund, formed from the same action-verb 'resign'. But (7) contains gerunds of two different action-verbs, so cannot be decided in the same simple way. Brand's explanation, that the baton-lowering and the signalling could occupy different regions in a world where the signalling convention were different, only holds provided that the operator *Dthat* is applied to the singular terms *properly contained* in each side of the identity. But each side of the identity is itself a description, or is equivalent to one, so why should the application stop short of the whole term? Brand says, 'let α^+ be the result of applying the *Dthat* operator to α' (Brand, *op. cit.*, 148), which implies that the operator applies to the description as a whole, in case α is itself a description. Brand's way with (7) shows that he stops short. There is, it turns out, a good reason why he must do so: if he does not, the criterion (5) becomes ineffective. For someone who insists that a sentence like (7) is true, in just the same way that (3) is true, the application of *Dthat* to each side as a whole creates a necessary identity, which means that (5) is satisfied while (7) is true, but had we relied on (5) to tell us whether (7) is true, since the truth of the sentence of the form $e^+ = f^+$ waited upon the truth of that of the form $e = f$, and vice versa, the criterion is powerless to decide the identity it was intended to decide. Brand can of course claim that this is a good reason to limit the application of *Dthat* to terms properly contained in the terms flanking the identity. But the friend of Davidson will claim that this is mere prejudice, and has no independent support.

One of the more appealing aspects of the particularist view of events is that it allows true identities concerning events, in particular actions, which are described in different and logically independent ways, without prejudice to the identity of the events concerned, just as we can do the same for material objects. As he uses it, Brand's criterion requires that *whenever* events are described using logically independent nouns, they will be different. This cuts down radically the number of true contingent identities his theory allows, and radically multiplies actions. It brings his theory much closer to that of the property-exemplification theorists than to that of Davidson. He has to deny not only disputed identities such as (7), but also identities such as

(8) David's falling off a horse at age five = the most unforgettable event in David's life

which even some property-exemplification theorists would accept as true (see Jaegwon Kim, 'Events as property exemplifications', in Brand and Walton (eds.), *op. cit.*, pp. 159–177, esp. p. 168.)

The problems with Brand's criterion suggest that we redirect attention to Davidson's proposal, which has at least the advantage that it does not involve modal notions. It remains true, however, that the particularist theory of events has no undisputed general criterion for event identity. This puts it about level with the theory of material objects upon which it is modelled.

University of Salzburg

© PETER M. SIMONS 1981

RUSSELL ON THE REDUCTION OF PARTICULARS

By ALBERT CASULLO

THE dichotomy between universals and particulars is one of the most fundamental in metaphysics. This dichotomy naturally gives rise to the most common sort of metaphysical question: Is the dichotomy irreducible or is there some way of overcoming the apparent dualism? The typical approach to overcoming the dualism has been to deny that there are universals. Very few philosophers have argued that particulars are reducible to universals. The most sustained effort in this latter direction appears in the later writings of Bertrand Russell. D. M. Armstrong has recently argued that Russell's efforts were not successful (*Nominalism and Realism,* Cambridge 1978, chapter 9). The primary purpose of this paper is to develop Russell's theory and defend it against Armstrong's objections.

I

Russell's primary motivation for arguing that particulars are reducible is epistemological. He held that an ontology which admits universals has only two alternatives with respect to particulars: (1) analyse them as complexes of qualities, or (2) admit that there are bare substrata. An ontology admitting such entities faces at least two serious problems. Although qualities are experienced, the substrata in which they allegedly inhere are not. Therefore, some justification must be offered for the belief that there are such entities. Furthermore, the existence of such entities allows for the possibility of distinct particulars with all qualities in common. Since substrata are indistinguishable apart from their qualities, two substrata having all qualities in common cannot be distinguished. Therefore, one cannot know that two particulars having all qualities in common are distinct and counting becomes theoretically impossible.

The primary obstacle to an ontology without substrata, according to Russell, is accounting for space-time order. A necessary condition for such an order is that there be at least some spatio-temporal relations which are asymmetric and transitive. Since the same quality can occur in a number of different places or at a number of different times, qualities cannot serve as the terms of such relations. For example, if *A* wholly precedes *B*, where *A* and *B* are qualities, it does not follow that *A* and *B* are not identical. Therefore, it appears that only substrata can stand in spatio-temporal relations.

Russell attempts to overcome this difficulty by introducing a relation called "compresence" which holds between two or more qualities. Using this relation he constructs groups of qualities, called "complete complexes of compresence," which have the two following properties: (1) all members of the group are compresent; (2) given anything not a member of the group, there is at least one member of the group with which it is not compresent. Complete complexes of compresence can serve as the basis for generating space-time order since it is empirically improbable, although not logically impossible, that two such complexes repeat themselves.

II

Armstrong attacks Russell's account on two fronts. First, he argues that it is committed to the necessary truth of the principle of the identity of indiscernibles (PII). He then presents three internal criticisms of the account which are independent of its alleged commitment to the PII. The argument based on the PII will be considered first.

Armstrong begins by distinguishing between a strong and a weak form of the PII. According to the strong form, different particulars must differ in at least some nonrelational properties; while, according to the weak form, different particulars must differ in at least some properties and these may include relational properties. Against the strong form of the principle, Armstrong argues: (1) that it is not a necessary truth; and (2) that there are scientific grounds for thinking that it is not true at all. Against the weak form of the principle, he argues: (1) that it involves the Bundle Theory in circularity; and (2) that it is not a necessary truth.

The first question which arises is whether Russell is committed to the *necessary* truth of the PII in either of its forms. Russell agrees that the truth of the PII is a logical consequence of the view that particulars are complexes of universals. But he is also emphatic in maintaining that it is a fortunate empirical fact that there are complexes of qualities which do not recur; i.e., which do not have all qualities in common. Because of this fortunate feature of our world a complete description of it, including its spatio-temporal framework, can be provided solely in terms of qualities. Since Russell did not hold that it is a necessary truth that objects in the actual world are nothing but complexes of qualities, he is not committed to the *necessary* truth of the PII. His position is clearly summarized in the following passage:

> To suppose that a complete complex can recur is to suppose that a complete complex can have the relation of earlier-and-later to itself. This, I assume, does not happen, or at any rate does not happen within any ordinary period of time. I do not mean to deny dogmatically that

> history may be cyclic, as some Stoics thought, but the possibility is too remote to be taken into account.[1]

The next issue to be addressed is which form of the PII Russell embraces. Since he holds that relations do not individuate, Russell agrees with Armstrong that the bundle theorist must appeal to the strong form of the PII. Armstrong's rejection of this form of the PII is based on the claim that it cannot be defended by appealing to differences in the spatio-temporal position of otherwise similar objects, since such a difference is not a difference in monadic properties. Russell, however, rejects this claim in the case of physical space-time as well as visual space. He holds that position in the visual field is absolute. Whatever is present in the centre of one's visual field has a quality called "centrality." Similarly, whatever is to the right is "dexter," whatever is to the left is "sinister," whatever above "superior," and whatever below "inferior" (*Human Knowledge*, pp. 298–99). It should be emphasized that these are qualities of the visual field and not *relations*. These qualities account for the difference between complexes in the visual field that appear to have all properties in common. In the case of physical space-time, Russell argues that complete complexes of compresence, which are complexes of monadic properties, have the necessary formal properties for constructing space-time order. Hence, he concludes that a complete complex of compresence 'counts as a space-time point-instant' (ibid., p. 304). Once physical space-time has been constructed in terms of complete complexes of compresence, one can appeal to an object's spatio-temporal position in defending the truth of the strong form of the PII.

III

Armstrong's three internal criticisms rest heavily on his claim that Russell identifies ordinary particulars with complete complexes of compresence. Russell, however, states explicitly that complete complexes are space-time point-instants rather than ordinary particulars. Furthermore, he says very little about the analysis of ordinary particulars. This is due to his belief that relativity theory has shown that the physical world is a four dimensional manifold of events rather than a manifold of persistent, moving objects. What were formerly thought to be persistent pieces of matter are in fact series of events in neighbouring regions of space-time related by certain physical laws. Hence, Russell devotes most of his efforts to shewing that events and space-time order can be constructed from qualities. His views about ordinary particulars must be inferred from his views about events and space-time order.

---

[1] Bertrand Russell, *Human Knowledge: Its Scope and Limits* (New York, 1948), p. 306. In fairness to Armstrong, it should be pointed out that he acknowledges that Russell does not always regard himself as committed to the necessity of the PII. Russell does not help matters when he slides from 'The identity of indiscernibles, which follows analytically from our theory, . . . ' to 'Indeed, I should claim it as the principal merit of the theory I am advocating that it makes the identity of indiscernibles analytic.' See *An Inquiry into Meaning and Truth* (London, 1940), pp. 102–103.

What then is the relationship between ordinary particulars and complete complexes? It appears that an ordinary particular is a series of events standing in certain causal relations. Events, however, are analysed as incomplete complexes of compresence which occupy a continuous region in space-time. An event is part of every complete complex which is a space-time point at which it exists. Therefore, a temporal stage of an ordinary particular is an incomplete complex of compresence while an ordinary particular is a series of incomplete complexes related by causal laws. A complete complex of compresence is neither an ordinary particular nor a temporal stage of an ordinary particular.

We will now turn to the three internal criticisms. The first begins with the premise: If something is a particular if and only if it is a complete complex of compresent properties, then a complex of compresent properties which is not complete is not a particular. Now consider a particular having the set of properties $\{x_1 \ldots x_n\}$. 'Might there not be another particular having no properties except ones which the original particular possessed, but which lacked some of the properties possessed by the first particular—its visual properties, say?' (*Nominalism and Realism*, p. 97). The theory must deny the existence of this particular because it would not be a *complete* complex of compresent properties.

This argument is difficult to interpret since it is not clear whether Armstrong intends to include among the properties of the original particular its spatio-temporal properties. If they are included, it appears that there could not be another particular having no properties except ones which the original possessed since that would result in two different particulars occupying the same place at the same time.[2] On the other hand, if the spatio-temporal properties of the original particular are not included among its properties, then there could exist another particular having no properties except ones which the original possessed provided that it differed from the original in its spatio-temporal location. Can Russell's theory accommodate such a particular? Recall that for Russell a complete complex of compresence is a space-time point-instant while an ordinary particular is a series of incomplete complexes. Consequently, a particular consisting of a subset of qualities of the original particular would exist if two conditions were satisfied: (1) there is an incomplete complex consisting of the subset of qualities of the original particular; and (2) this incomplete complex is part of a complete complex of compresence different from that to which the original set of qualities belongs.

---

[2] Armstrong allows this possibility for what he calls "abstract" particulars but not for "concrete" particulars. See *Nominalism and Realism*, pp. 120–121. Aside from the fact that it is questionable whether the notion of an abstract particular is a coherent one, Russell is concerned only with ordinary or concrete particulars.

Since qualities can occur in many different complexes, there is no reason why there could not be two such complete complexes. From this it does *not* follow that every subset of qualities of the original particular constitutes a particular. For it is a purely contingent matter which subsets of qualities of the original particular are parts of other complete complexes.

Although Armstrong's first internal criticism fails, the premise that something is a particular if and only if it is a complete complex of compresence must be rejected. Consider the case of a person perceiving two cups. The perceived state of affairs involves two different ordinary particulars yet there is only one complete complex of compresence. Therefore, a complete complex can involve more than one ordinary particular. Russell recognized this:

> We now have to consider what parts of *W* [a complete complex of compresence] can be wholes requiring names for the expression of judgments of perception. The total *W* may be analysed into a number of qualities but this analysis will not, by itself, enable us to explain such judgments of perception as "*A* is to the left of *B*". These demand the analysis of *W* into what we should be tempted to call "substantial" rather than "conceptual" parts ... We now require a relation of "overlapping", which plays a part in the construction of perceptual space analogous to that played by compresence in private time (*Inquiry*, p. 338).

Several important points emerge in this passage. First, that within a complete complex of compresence there are smaller complexes, or substantial parts, consisting of a subset of the qualities of the complete complex structured by the relation of *overlapping*. Second, a complete complex of compresence can contain more than one instance of *overlapping* and, hence, more than one substantial part. Finally, in the case of perceptual judgments, such as '*A* is to the left of *B*', involving two particulars within a single complete complex of compresence, '*A*' and '*B*' refer to substantial parts of the complete complex. All of this strongly suggests that it is the substantial parts of the complete complex, rather than the complete complex itself, which takes the place of ordinary particulars in Russell's theory. Armstrong overlooks this crucial feature of the theory. Moreover, once it is recognized, the remaining criticisms can also be answered.

According to the second internal criticism, Russell's theory cannot allow for the following possible state of affairs. Consider three particulars, *a*, *b*, and *c*, with the following characteristics: *a* has among its properties *P* and *Q* but lacks R; *b* has *Q* and R but lacks *P*; *c* has *P* and R but lacks *Q*. Furthermore, suppose that it is nomically impossible that any particular have all the properties *P*, *Q*, and R. Since *P* is compresent with *Q*, *Q* with R, and *P* with R, they form a complex of compresence.

Therefore, on Russell's theory, there exists a particular which has all the properties *P*, *Q*, and R. But, by hypothesis, such a particular cannot exist.

Armstrong's argument involves two assumptions: (1) if *P* is compresent with *Q*, *Q* with R, and *P* with R, then *P*, *Q*, and R form a complex of compresence; (2) if *P*, *Q*, and R form a complex of compresence, then there exists an ordinary particular which has the properties *P*, *Q*, and R. If either is rejected, the argument fails. In *The Structure of Appearance* (3rd ed., Dordrecht 1977, p. 147) Nelson Goodman presents an argument strikingly similar to Armstrong's but maintains that what it establishes is that it is wrong to suppose 'that for three or more qualities to be together [i.e., compresent] means simply for each two of them to be together.' He goes on to suggest that togetherness be construed as a relation which obtains not merely between individual qualities but also between sums of qualities. One then can hold that three or more qualities form a complex of compresence if and only if every sum of one or more of the qualities is compresent with the sum of the rest. Thus, (1) can be rejected.[3] (2) is also questionable. For, as we saw earlier, Russell holds that a complex of compresence may contain substantial parts and it is the substantial parts of such complexes which take the place of ordinary particulars in his account. Since it does not follow from the premise that *P*, *Q*, and R form a complex of compresence that there is a substantial part of the complex which is *P*, *Q*, and R, neither does it follow from this premise that some ordinary particular has all the properties *P*, *Q*, and R.

Armstrong's final argument begins with the observation that many properties of particulars involve essential reference to the proper parts of these particulars. Since the parts are themselves particulars, some properties involve the notion of further particulars. The bundle theorist, of course, will maintain that these further particulars are also bundles of properties. But the new bundles may include properties which refer to further parts. Therefore,

> it is at least logically possible that this process should go on *ad infinitum*. A particular may lack any ultimate parts. But for such particulars, it is suggested, it is impossible to carry out the resolution of particulars into bundles of properties (ibid., p. 100).

Armstrong does not explicate the notion of a proper part in the context of the discussion but spatial parts are offered as an example. Consequently, our remarks will be restricted to spatial parts. The assumption that the spatial parts of a particular are themselves particulars is hardly an uncontroversial ontological claim. It has been suggested, for example, that talk about parts of an object can be construed in terms of adverbial modifiers of verbs which are true of particulars as a whole (Wilfrid Sellars, 'The Adverbial Theory of the Objects of Sensation,' *Metaphilosophy* 6, 1976, pp. 146–147). But even if we grant that proper parts of particulars are genuine particulars, this final argument falls short of its mark for two reasons. First, Russell's theory is rich enough to provide an account of the spatial parts of a particular. For a complete complex of compresence often contains smaller complexes formed by the relation of *overlapping* and these smaller complexes stand in spatial relations. Second, the regress in question is not vicious. Nothing in the argument shows that the bundle theorist must eventually admit particulars which cannot be eliminated in favour of properties by further analysis. It simply points to the truism that if an ordinary particular consists of an infinite number of other particulars then, according to the bundle theory, such a particular must consist of an infinite number of bundles of properties. Hence, the analysis of such a particular into bundles of properties cannot be completed. But, of course, any alternative account of ordinary particulars will be open to the very same objection. If the parts of a particular are also particulars and an ordinary particular has no ultimate parts, then no analysis of it can be completed.

---

[3] For an alternative approach, see A. J. Ayer, *Russell and Moore: The Analytical Heritage* (London, 1971), pp. 75–76.

We can now see that Russell's approach to reducing particulars is more promising than might initially appear. It is not open to the standard objection that any such theory is committed to the necessary truth of the PII. One must also recognize that the central idea of the theory is not the replacement of ordinary particulars with complete complexes of compresence. Instead, it begins at a more fundamental level with the construction of space-time order. In conclusion, I want to suggest that if the theory ultimately fails it will be the result of inadequacies in its treatment of space and time rather than its account of ordinary particulars.[4]

The University of Nebraska—Lincoln

© ALBERT CASULLO 1981

---

[4] I want to thank Robert Audi, Philip Hugly and an anonymous referee for their helpful comments, and the Research Council of the University of Nebraska for its financial support. An earlier version of this paper was presented at the American Philosophical Association, Western Division Meetings, April 25–27, 1980.

HESS ON REASONS AND CAUSES

By PETER SMITH

ONE of the many insights of Davidson's classic paper 'Actions, Reasons, and Causes' was this: a singular causal truth and the corresponding causal law which backs it need not trade in the same family of concepts. Thus, to borrow a well-worn example, suppose that the event reported on page 5 of the *Times* (namely Monday's hurricane) caused the event reported on page 13 of the *Tribune* (namely Tuesday's bridge collapse). It would obviously be insane to suppose that the truth of the singular causal claim needs to be warranted by a back-up law couched in terms of 'page 5 reports' and so on; the property of being reported on page 5 is evidently not causally salient. And, as Davidson remarks, it would be scarcely less ridiculous to suppose that there are supporting laws which speak of 'hurricanes' and such things as 'bridge-failures'. Of course, the property of being a hurricane is causally salient in a way that the property of being reported is not, for an event counts as a hurricane in virtue of its other, law-engaging, physical properties: but still, it is these other properties (pressure gradients and the like) which will no doubt be mentioned in a causal law backing up a singular causal claim about a given hurricane—the concept of a hurricane itself is too coarse to feature in the grounding laws. So, to repeat, a singular causal truth and its supporting law need not engage the same set of concepts. Therefore—to get to the key point—the explanation of an action by reference to a man's beliefs and desires may yet be causal, even if there are no back-up laws framed in psychological terms. For the backing law for a psychological explanation may be couched in quite other terms, perhaps physicalistic ones.

All this is by now excessively familiar, and would hardly bear repetition were it not for the fact that these points have recently been badly misunderstood by Peter Hess in his paper 'Actions, Reasons, and Humean Causes' (ANALYSIS 41.2, March 1981 pp. 77–81). Hess wants to show that Davidson's handling of his points in fact gives no aid or comfort to the defender of the traditional commonsensical view that reasons are causes. On the contrary (Hess argues), Davidson takes the ground from under the traditionalist's feet, leaving us with a position that no self-respecting defender of the old view could recognize as his own. Let me explain.

Consider again the statement that the event reported on p. 5 of the *Times* caused the event reported on p. 13 of the *Tribune*. Hess regards this as Davidson's paradigm of a heteronomic singular claim (i.e. the sort of singular causal claim which is warranted, if at all, by a law couched in quite distinct terms). Now, one feature of this supposed paradigm is that it identifies the events it asserts to be causally related via properties which are quite independent of what gives the events their causal efficacy. Hess seizes on this feature, and takes it to be shared in Davidson's view by the singular claim that (the onset of) a given belief or desire caused a particular action. Here too we have a heteronomic claim which again identifies the events which it asserts to be causally related via properties which are quite independent of what gives the events their causal efficacy. In other words, Hess thinks that in Davidson's view the property of being (the onset of) a belief or desire is no more a causally salient property of events than the property of being reported on page 5. Or in Hess's phrase, Davidson's position 'reduces primary reasons to the status of epiphenomena'.

Obviously, if this were Davidson's position, then it would indeed be an illusion to suppose that he was reformulating anything like a common-sensical causal theory of action. For while no-one assumes that the event mentioned on page 5 caused a bridge to collapse by virtue of being so mentioned, a traditional theory of action does suppose that the event of coming to desire a beer causes one to open a can by virtue of being that desire. As Hess puts it, advocates of the common-sense position 'will certainly not accept the view that the causal efficacy of whatever causes an action is in no way affected by the presence or absence of primary reasons'.

But Hess has misconstrued Davidson. A heteronomic singular claim *may* be couched in causally irrelevant terms (as in the overworked newspaper example), but there is no necessity about this. The claim that Monday's hurricane caused Tuesday's bridge failure is again heteronomic, as Davidson insists, yet it identifies what produced the failure in causally salient terms (and of course it makes some sense to say that the hurricane caused the collapse by virtue of being a hurricane). So Davidson's view, that a singular claim relating reason and action is a heteronomic causal statement, leaves wide open the option of holding that the notions of belief and desire are (as the traditionalist thought) causally salient. And this option is the one that Davidson embraces. In brief, the concepts of belief and desire are more like the concept of a hurricane than that of a reported event—causally salient even if not fit for appearing in hard-edged causal laws.

Hess, therefore, has badly misread the position in 'Actions, Reasons and Causes'. He treats the newspaper reports example as a general paradigm of heteronomic claims, ignoring Davidson's own examples of other such claims that lack the feature of employing causally irrelevant concepts (the feature on which Hess places such weight). The mistake here is compounded by a misreading of the position of 'Mental Events', to which Hess appeals for further support of his own argument. This additional mistake is perhaps also worth exposing.

Hess offers the following definition of what he means by 'epiphenomenal':

> A property P is epiphenomenal with respect to the relationship between an event C and its effect E iff
> (i) P is a property of C;
> (ii) It is not the case that C would not have caused E had it not had property P.

He then adds:

> If the theory of anomalous monism is true, then every mental property is epiphenomenal with respect to every mental event and its physical consequences. If there were even one mental property that is not epiphenomenal in that sense, then there would be at least one psychophysical law. But, according to Davidson, there aren't any such laws.

So once more it apparently follows that mental events are like newsworthy events in respect of their causal potency. Just as the event reported on page 5 does not cause the bridge failure by virtue of the epiphenomenal property of being so reported, so the event which is (the onset of) a belief or desire does not cause the ensuing action by virtue of being a belief or desire. Again we are landed with a quite untraditional view.

But is Hess right in thinking that the doctrine of anomalous monism renders all mental properties epiphenomenal in his sense? Remember that the doctrine is quite compatible with the claim that mental properties are supervenient on (law-engaging) physical properties, in the sense that there cannot be two events alike in all such physical respects but differing in some mental respect. Or at least, Davidson plausibly argues that his monism is consistent with the supervenience thesis (see p. 214 of the reprint of 'Mental Events' in *Essays on Actions and Events*), and Hess offers us no rebuttal. Suppose therefore that we accept the supervenience thesis. And let C be, say, the physical event which is the onset of Jack's desire for a beer, E the consequent event of Jack's opening the can, and P the mental property of being the onset of a desire for beer. Then is property P epiphenomenal?

Well, by the supervenience assumption, C cannot lack property P while remaining the same in all law-engaging physical respects (contrast the case of the event which stays physically the same whether or not it gets reported on page 5). So, an event which lacks P must be an event with some different basic physical characteristics from those actually possessed by C, and on a plausible view about event identity, this will necessarily be a distinct individual event. But if that is right, then it will be vacuously true that C would not have caused E had it not had property P (for it would not then have existed, and so couldn't cause anything!). Thus, by the definition, the supervenient property P is *not* epiphenomenal, and Hess's argument does not go through.

If this counter-argument is resisted on the basis of an alternative view about event-identity, and I grant that the issue is a disputable one, it will still be to no avail. For we can retreat to a simpler argument by dilemma. Either the supervenient mental property P is or is not epiphenomenal. If it isn't, then Hess's position collapses. If it is, then the property of being a hurricane, which equally supervenes on an event's basic law-engaging physical characteristics, will by parity of reasoning also be epiphenomenal. But any theory which can place mental states and hurricanes on a par as causal factors (even if they are dubbed 'epiphenomena') will satisfy the most ardent traditionalist, and so again Hess fails to drive a wedge between Davidson and the traditional theorist.

Davidson's monistic causal theory of action allows beliefs and desires to be both causally salient and supervenient on the physical. It is only by overlooking these two familiar points that Hess can propound his jejune arguments.

University College of Wales, Aberystwyth © Peter Smith 1981

CONDITIONAL MURDERERS

By Christopher Bryant

IT has often been remarked that philosophical fitness is stunted by too meagre a diet of examples; things are worse still when the examples are undercooked as well. Considering the intellectual effort which has been devoted to trying to provide something called a semantics for "counterfactual" conditionals which is to be interestingly different from a parallel construction for other, "indicative", conditionals, it is surprising that more of the effort did not address the question whether there really is a logically important difference, particularly since powerful arguments have been made out on various occasions for the contrary claim (e.g., by Ayers in *Mind* 1965; Slater in *Philosophical Quarterly* 1974; Mackie in *Truth, Probability and Paradox*, p. 65). Unless they can be refuted, there can be no reason for exploring the supposedly different ways in which possible worlds may be encouraged to put their gross cosmic products at the service of a would-be user of 'if'. In fact the arguments have not been refuted so much as merely brushed aside, very largely on the strength of a single example of a pair of conditionals drawn from Adams (*Foundations of Language*, 1970). Thus when Lewis (*Counterfactuals* 1.7, p. 27) turns to the problem of what has gone wrong when a counterfactual turns out to have a true antecedent after all, he says that it is not safe merely to cancel the misleading implication by restoring the antecedent to an indicative mood, because the Adams example shows that counterfactual and indicative conditionals may differ in truth value. If the argument is to have any weight, the Adams example must be of a pair of conditionals which differ only in one respect—the grammar of the counterfactual version must imply that the antecedent is false. Lewis clearly thinks that this is true of the Adams example: when he first discusses it (p. 3) he says that it shows that there are two different *sorts* of conditional, not just one 'that can appear as conditional or as counterfactual depending on the speaker's opinion of the truth of the antecedent'.

The example in question is the pair

(1) If Oswald did not kill Kennedy, someone else did
(2) If Oswald had not killed Kennedy, someone else would have.

Nobody, it has seemed, can say that the truth conditions of these conditionals are the same: (1) commits its user at most to the existence of one actual assassin, of unknown identity, whereas (2) evidently trails with it a scenario involving a queue of at least two would-be attackers, ranged in ascending order of reliability. But, according to the separatists, (2) is just the counterfactual version of (1); hence it cannot be true, in virtue of this one example, that counterfactuals obey the same logical rules as other conditionals. So obvious is it that (1) and (2) are part of different stories that it is surprising that the separatists did not more carefully consider whether they are closely *enough* related to each other to discharge their appointed role. For if (2) is the counterfactual version not of (1) but of a quite different open conditional, then nothing whatever follows about the logical relations of counterfactuals as such. One cannot show that they are different from other conditionals by remarking that the truth conditions of (1) are different from those of, say, 'She would have left the room had not Mr Collins addressed her again'. This is simply a different conditional, and one can no more use its difference from (1) for the purpose than one could use the logical relations of 'Mr Collins addressed her' and 'She did not leave the room' to illustrate the logical effects of negation. The Adams example will work only if the two sentences differ solely in respect of whether it is given or not that Oswald killed Kennedy.

It is important to be clear about this because the recent exchanges in ANALYSIS between Lowe and Davis somewhat obscured it. In his first paper (39.3, June 1979, pp. 139–141) Lowe correctly observed that if an open ("indicative") version of (2) could be found which was not the same as (1), Lewis' argument would collapse, and he provided one in his

(7) If Oswald has not killed Kennedy, someone else will have

(similar points were made by Ellis in *Rational Belief Systems,* Oxford 1979, p. 67). Davis, replying for the separatists in 40.4, October 1980, pp. 184–6, missed the point:

> Once it is granted . . . that (1) and (2) are not equivalent, Lewis' conclusion follows. The possibility that some *other* conditional is equivalent to (2) is irrelevant . . Any dispute . . as to whether (1) really is the indicative counterpart to (2) is verbal and therefore pointless. The type of correspondence Lewis and company had in mind is precisely that between (1) and (2).

This was confused: as Lowe showed in his rejoinder (*ib.* p. 189), the question is whether the correspondence between (1) and (2) is what Lewis thought it was. Since Davis' other attempts to throw doubt on the relationship between (7) and (2) did not survive examination, Lowe concluded that Lewis' argument was still in danger, but he could not be sure that a better example might not be found, and the force of his arguments was dissipated in manoeuvres with the notion "material conditional" (it is high time this lethally misleading piece of terminology was extirpated from the philosophical vocabulary) and with notions from Lewis like "variably strict conditional" which become vacuous with the death of the Adams example.

A much stronger position is tenable: not only do (1) and (2) entirely fail to correspond in the required way, but it can also be shown that no examples could provide a pair corresponding in the way required but with different truth values.

The first thing to grasp is that not only has (2) a range of open equivalents (Lowe's (7) is one of several) but in addition (1), so far from being one of these, has a range of counterfactual versions of its own: these two ranges are entirely distinct. I do not dispute that (1) could be produced as an inaccurate anglicization of something with a hook in it, but it has a more straightforward role as a conditional which could be used in *modus ponens* inferences in English. The baffled police chief of Dallas, let us say, is confronted by a watertight alibi for Oswald, his hitherto most promising suspect. So (since Kennedy was certainly killed) he argues: 'If Oswald didn't kill Kennedy someone else did. Oswald didn't after all; so someone else must have, and I shall have to get busy finding out who'. The platitudinous nature of the first premise must not be allowed to conceal its role in the inference. Moreover the inference remains valid, and for the same reason, under consistent substitution of tenses. Thus a despairing secret service man, previously aware of Oswald's plans but now finding as the shot cracks out that, like Macavity, Oswald is absent from the scene of crime, may argue 'If Oswald isn't killing Kennedy, someone else is, so . . ', or even earlier (a fatalist, this fellow) 'If it turns out not to be the case that Oswald kills Kennedy then it will be someone else who does . . . ' (it is of course much less plausible that anyone should argue thus, but only because of the difficulty of establishing the premises—precognition?). The more complicated auxiliary verbs of the future tense version are conspicuous, and necessary because of the multifarious jobs auxiliary verbs, particularly 'will' and 'would', have to perform. 'If Oswald will not kill . . .' means if he refuses; 'If Oswald does not . . . someone else will' gives us the wrong temporal relation, for it implies that an alternative murderer will subsequently spring into action following Oswald's failure or omission. The tenses, as Prior has shown, function in two ways—as indexicals, and as terms in what he called a theory of the earlier-later relation. So far as they do the first job, consistent substitution can no more affect validity than moving from 'That is mine' to 'This is yours' in the mouths of respective interlocutors; but the substitution must not affect the earlier-later relations, ambiguously expressed though these may be. The absolute tenses make no logical difference, and the conditional concerned in each of the above pieces of reasoning is one and the same.

Now each of these versions has a counterfactual correspondent, namely 'If Oswald had not killed Kennedy, it would have had to be someone else who did', 'If Oswald were not killing Kennedy, it would have to be someone else who is', and 'If it were not to be Oswald who will kill Kennedy, it would have to be someone else who will'. The modal-looking auxiliary verbs are of course not part of mysteriously modal consequents but part of the conditional operators. As for the periphrases (' . . . would have to be . . . who . . '), these are necessary to avoid the implications of allowing 'would' with an active subject to get at the main verb directly. 'If Oswald were not killing Kennedy, someone else would be' broadcasts motive and purpose too widely. 'If Oswald were not killing Kennedy, someone else would' in addition produces the different temporal structure charted above.

I can be brief with the other half of the story: Lowe has provided the beginnings of it already in his (7), and there are the others too, 'If Oswald did not kill Kennedy, then someone else was to', and, obviously, 'If Oswald does not kill Kennedy then someone else will'.

It thus turns out that (1) and (2) do not do what the separatist needs. They differ exactly where everybody can see that they differ, in being different conditionals, from different stories; the respective parts of the separate stories can be rejected or entertained by the utterer independently. The same will be true of other pairs of examples that might be used for such a purpose: only lack of care with auxiliary verbs can make them look plausible.

Davis advances two close relations of the Adams pair as further examples to separate open conditionals from counterfactuals:

(5) If Oswald did not kill Kennedy, no one did
(6) If Oswald had not killed Kennedy, no one would have.

He says flatly that (5) is false and (6) true. This assertion is followed by some possible-world claims which accord with the alleged truth values, but they cannot be Davis' reason for ascribing the values or his argument would be circular—the truth values are supposed to show why his particular line of possible worlds is a good market proposition. In fact (5) is true: as Lowe says, its gist is that Oswald was the only possible killer. To be what Davis needs, (6) must say that Oswald, who did in fact kill Kennedy, was the only person who can have been the killer, and this has just the same truth value as (5) and for the same reason. It can appear different only if the ambiguous 'would' is allowed to suggest (through a buried future tense) that the total absence of assassination would be somehow consequent upon Oswald's abstention; but again this is a different conditional whose indicative equivalent is 'If Oswald did not kill Kennedy then no one would', or, in the present, 'While Oswald does not kill Kennedy no one will'.

It is also possible to be fairly brief about whether an open/counterfactual pair *could* display the required difference in truth value. It would have to be the case that the usefulness of the open member turned on whether the antecedent was actually true or false (logic could scarcely absorb the consequences of suggesting that whether the speaker *thought* it was true is important). In other words, the reliability of the inference rule 'given *p* you may infer *q*' would depend on whether you are given *p*. But this would be absurd: if you are not given *p* you cannot use the rule anyway. Moreover if the actual falsity of the antecedent threatened the conditional, then no *modus tollens* inference could be valid, which is also absurd. It is less obvious that being in a position to use a counterfactual is *eo ipso* to be in a position to use the corresponding open conditional if the antecedent turns out after all not to be certainly false. An open conditional fits into the language of argument and discovery, as a vehicle for carrying you from what is or may be known to the presently unknown. While the antecedent is currently taken to be false no such role remains for it, because the conditional cannot be used in practice in a *modus ponens* inference, but neither can it figure in a *modus tollens* argument, for by hypothesis the conclusion of such a putative argument is already taken to be known. (This does not of course mean that someone for whom the conditional is counterfactual cannot consider the validity of an argument produced by someone else for whom it was open.) The counterfactual fits into the language rather of description and explanation, of assessment. It follows that if it turns out that the antecedent may after all be true the narrative will have to be reconsidered, as a whole or in part; but it does not follow that the previously counterfactual conditionals with which it was first constructed will have to be discarded rather than merely returned to store. We explode a counterfactual itself only by discovering that its consequent was or could have been false while the antecedent was true, which is precisely what would settle the hash of the equivalent open conditional.

Thus I may use the ambiguous conditional 'If the princes had disappeared under mysterious circumstances their mother the Dowager Queen would never have accepted Richard's pension and a place at court' either as part of an argument to show by *modus tollens* that he could not have done away with the princes, or, if I take it to be established already that he did not, as part of an oblique description of Elizabeth's fiercely maternal character, as a way of saying for its own sake that she was not that sort of woman. But this is just what the *modus tollens* argument depends on. To demolish the conditional in either case you would have to show that she was or at least could well have been that kind of woman. You could do so, of course, by showing that she actually did swallow the actual taking-off of her sons, or, more generally, by showing that nothing about her known career, or about feminine regal psychology, entitles me to count on her finer feelings. This shows what I should have to do to justify either the argument or the character sketch. I cannot however use the more direct and less general evidence to *establish* my conditional in either case. Even if I could take myself to know that (a) Richard did have the princes killed, and (b) Elizabeth did not take the pension, I have no business yet with a conditional; what I have so far is a conjunction, and to move it into conditional form would be to dive directly into a *post hoc* fallacy. For you one combination of the component truth values is sufficient, but not necessary; for me no combination is even sufficient. This consideration spells death for the whole enterprise of other-worldly semantics for conditionals, something which could come as a surprise only to those who regard conditionals as truth-functional. But the mere truth or falsity of (a) cannot open a divergence between counterfactual and open versions.

What may indeed come as a surprise is that not only do the Adams pair not correspond in the required way, they do not even differ in truth value in the way they have been supposed to. Discussions of the example have invariably begun by postulating that reality is the Warren word made flesh: Oswald is to have killed Kennedy acting by his wild lone, and there are to be no other plausible suspects. We can indeed infer from this that (2) is false, for it multiplies the murderous. But then (1) must be false too, for what it says is that if Oswald did not do it (after all), then there is an alternative culprit. If we accept the report, (1) needs to be expressed in counterfactual terms (possibly part of the reason why separatists have mishandled it, and a further indication of why looking for an Adams-type pair is futile); but ignoring this, if its antecedent were after all true something would have to be wrong with the report, at least where it says Oswald killed Kennedy. But then knowing just this much, all the antecedent of (1) mentions, could we reasonably infer that someone else killed Kennedy, rather than, say, that Kennedy was not killed at all? No. This is not a matter for some mythical "linguistic intuition" (as dallied with by, e.g., Lewis (p. 28), Ellis (p. 76)) but for straightforward historical reasoning. Knowing just that this one part of the report is not true we cannot yet safely infer anything about what else in it remains true nonetheless. Let the *cognoscenti* swear that facts can be clearly descried in some other universe; no such facts could be used to support the inference on pain of vicious circularity.

So if (2) is false because there was actually no other would-be killer, then (1) cannot be true; equally, if it is on the cards that the killer of Kennedy may not have been Oswald, i.e. if (1) is true, then we cannot after all say that (2) is false. Those who desire to separate here one category of objects from another should inspect their muttons in a less casually Polyphemian fashion.

University of St Andrews © CHRISTOPHER BRYANT 1981

BUTLER'S PROBLEM AGAIN

By Eric Stiffler

IN his 'Killing People Intentionally, By Chance' (Analysis 41.3, June 1981 p. 156), Kim Davies proposes yet another solution to R. J. Butler's Analysis Problem No. 16 (37.3, March 1977, p. 97). Butler's problem, remember, was this:

> If Brown in an ordinary game of dice hopes to throw a six and does so, we do not say that he threw the six intentionally. On the other hand if Brown puts one live cartridge into a six-chambered revolver, spins the chamber as he aims it at Smith and pulls the trigger hoping to kill Smith, we would say if he succeeded that he had killed Smith intentionally. How can this be so, since in both cases the probability of the desired result is the same?

Davies' strategy is to focus attention on the act types involved in the two cases, throwing a six in an ordinary game of dice and killing Smith. He defends our intuitive ascriptions of intentionality by arguing, in effect, that only the latter can be instantiated intentionally. If this were the case, it would solve the problem—but I do not believe that it is.

Davies' solution turns on the notion of general intention: 'killing Smith is a consequence of hitting Smith with the bullet, a consequence which, unlike throwing a six, is the object of a general intention. That is, the scope of this intention is not limited to a particular occasion.' Brown has a general intention to kill Smith if he intends to kill him not just in the game of Russian Roulette but by any suitable method, the game providing but one suitable opportunity. It is consistent with the facts of the case to suppose that Brown's intention is general, and this, Davies suggests, is the reason we are inclined to regard the killing as intentional in spite of the unusual circumstances. Still, we cannot say it was intentional that *on this particular occasion* Smith be killed because in the game described Brown has no control over whether he kills Smith. However, if Brown meant to kill Smith by any means available, on this particular occasion it was intentional that Smith be killed.

So far I agree. We can imagine that prior to entering the game Brown decided to kill Smith. If his resolve was such that he would have killed Smith at the first opportunity, and this just happened to be the game of Russian Roulette, the killing is intentional because Smith's dying in these circumstances realizes Brown's general intention. But it seems to me that the same can be said of Brown's throwing the six.

We are told that Brown hopes to throw a six but the actual content of his hope can reasonably be viewed as more complex than this. What

Brown hopes to do is throw a six in the circumstances of an ordinary dice game. The fact that in an ordinary dice game a person has no control over whether he throws a six counts against regarding Brown's successful throw as intentional. But this is just to say that, as in the shooting case, it is not intentional that *on the particular occasion in question* the desired result occur. If Brown's hope is merely that on this particular toss of the die a six turn up, his throwing the six is not intentional. However, his hope need not be limited to a particular occasion. Just as the shooting case allows us to suppose that Brown has a general intention to kill Smith, the die-throwing situation permits the assumption that Brown has a general desire to throw a six. For example, Brown may hope to throw a six at least once in the course of the evening's play. The question is whether the object of his hope—to throw a six in an ordinary dice game at some time during the evening—can also be the object of an intention. If it can be, the symmetry between the shooting and die-throwing cases is restored. For notice that the hope expressed is not specific as to method or occasion of success. If it qualifies as an intention it would appear to be a general intention and so, by Davies' argument, on the occasion of the successful throw there is a sense in which it is intentional that a six turn up.

It is surely odd to say that anyone intends to throw a six in an ordinary dice game. The oddity can perhaps be explained in terms of the agent's lack of control over the outcome of his throw, but the notion of control employed here is murky. While partially cognitive, it also brings in the agent's causal powers and capacities. I think the oddity can be accounted for in purely cognitive terms. The reason I cannot intend to throw a six on a particular occasion of die-throwing in an ordinary game is that I have no idea how to go about doing it. Having some conception of how to perform an action seems to be a necessary condition for intending to perform it; at the very least, an agent must believe there are steps he can take to accomplish whatever it is he intends to accomplish. We risk equating intending with mere wanting unless some such constraint is placed on the actions a person intends to perform.

This proposal concerning the cognitive element in intending is based on the following view about the connection between intending and believing. In order for an agent X to intend to perform an action A, X must have a conception of how to bring A about; and X has such a conception if and only if X believes that there is an action (or series of actions) B, and a set of circumstances C, such that his doing B in C will lead or amount to his doing A. The connection is admittedly weak. It is not necessary that X have adopted or settled on a plan for doing A in order to intend to do A; nor is it necessary that X be familiar with or able to identify specific methods for bringing A about. This is in keeping with the phenomena of prospective intentionality. Brown can be said to have formed a prospective intention to kill Smith before he begins to consider ways and means; and he can prospectively intend to rebuild the carburettor on his car even though at the time he doesn't know the first thing about rebuilding carburettors. The deciding issue is whether the agent believes there are things he can do to bring about the end he desires. Lacking this confidence, it is difficult to see how X can seriously contemplate doing A let alone decide or resolve to do A. X may, of course, still *want* to do A when such a belief is absent, but it is doubtful we would allow that he intends to do A even assuming all the conative elements of intending are in place.

The implications of this partial analysis of intending for Butler's die-throwing case make clear its parallel to the shooting case. Brown cannot intend to throw a six in a particular dice game because, to the extent that he is rational, he will not believe there is any action he can take which will lead to a six turning up. He knows, of course, that by simply tossing the die he may throw a six; but he will not believe that simply tossing it will have this result. On the other hand, Brown can have the general intention of throwing a six in an ordinary dice game at some time during the evening or in the course of a series of ordinary games. There is nothing in the example to suggest he lacks the wants relevant for having this intention, and the cognitive requirement is clearly satisfied because, if he is rational, Brown will believe there is something he can do to throw a six in the context of a series of ordinary games, namely keep tossing the die. As a consequence, when a six finally appears we can say that its turning up was intentional on Brown's part since the event realizes his general intention to throw a six.

Our intuitive ascriptions of intentionality in Butler's cases are based on two assumptions: that in the die-throwing case Brown does not have a general intention like the one I have concocted, and that in the killing case Brown does have a general intention to kill Smith. There is nothing in the statement of the problem to warrant these assumptions. Perhaps we make them because we think a person would be likely to play the game of Russian Roulette Brown plays only if he had a general intention to kill Smith, and that a dice player usually bets on the outcome of a particular thow and not on that of a series of throws. But the cases themselves are symmetrical. They permit the attribution of sufficiently varied wants and beliefs to Brown to support all possible combinations of ascriptions of intentionality in the two cases.

Western Illinois University

© Eric Stiffler 1981

STEINER ON NOZICK ON THE RIGHT TO ENFORCE

By Patrick Wilson

IN *Anarchy, State and Utopia* (Oxford, 1974), Robert Nozick argues that if I have a right to your doing something, this does not entail that I have a right to force you to do it (pp. 90–95). For example, if you promise me not to murder someone, this does not entail that I now have a right (because of your promise) to force you not to do so, since this is a right which I possessed already, Nozick says, before you made your promise. Or, to take another example, if I promise not to force you to do A if you promise to do A, then your subsequent promise to do A cannot plausibly be regarded as entailing that I now have a right to force you to do A, since I have already promised not to force you. Quite apart from general considerations, these two cases at least, Nozick implies (p. 92), seem to exemplify the truth of his general thesis that rights do not entail further rights of right-enforcement.

Hillel Steiner (Analysis 41.1, January 1981, p. 50) argues that in the latter case 'Nozick's argument is self-contradictory since it implies at once that', in promising not to force you to do A if you promise to do A, I both am and yet cannot be 'under an obligation not to force'. On Nozick's general thesis, my having a right to your doing A (because of your promise to do it) does not entail my having a right to force you to do A. Likewise, Steiner points out, *your* having a right not to be forced by me to do A (because of my promise not to force you) would not, according to Nozick, entail *your* having a right to force me not to force you to do A. Therefore, Steiner goes on, since you would lack the right to force me not to force you to do A, it follows that I would lack the obligation not to force you to do A, since, he claims, this obligation is the correlative of the right which you would lack. Thus, Steiner concludes, if Nozick's general thesis is correct, then it is logically impossible for me to promise not to force you to do A, since this promise incurs for me exactly the obligation which, according to Nozick's general thesis, I must necessarily lack. Hence, Nozick's assumption that that promise is logically possible contradicts his general thesis.

Steiner's argument seems incorrect. My obligation correlative to your right to force me to keep my promise (not to force you) is not the obligation not to force you to keep *your* promise (to do A). It is the obligation not to force you not to force me to keep *my* promise (not to force you), since such force would be a denial of your right to force me to keep my promise. Hence, since (as Steiner points out) this is the right which you would lack, according to Nozick's thesis, then it follows that the correlative obligation which I would necessarily lack is the obligation not to force you not to force me to keep my promise (not to force you). I would not, as Steiner claims, necessarily lack the obligation not to force you to keep your promise (to do A). Hence it would not be logically impossible for me to incur an obligation not to force you to keep your promise (e.g., as in Nozick's example, by promising not to force you), since I would not necessarily lack this obligation.

The same result can be arrived at by starting with Nozick's own definition of rights as 'permissions to do something and obligations on others not to interfere' (p. 92). We can pass over Nozick's questionable use here of the term 'permissions'. (A better word might be 'entitlements'.) The important point in the present context is that, if you lack a Nozickian right to do something, then the obligation which I lack is the obligation not to interfere with your doing it. In Nozick's example, the 'something' in question which you lack a right to do is, as Steiner points out, to force me to keep my promise. The obligation, then, which I necessarily lack is the obligation not to interfere (by force, if necessary) so as to prevent your forcing me to keep my promise. It is not, as Steiner claims, the obligation not to interfere (by force, if necessary) so as to prevent you from keeping some promise of *yours*. Hence, since I would not necessarily lack an obligation not to force you to keep your promise, I could incur that obligation (by promising not to force you) without damage to Nozick's thesis.

One is left wondering, nevertheless, just how convincing Nozick's thesis really is. Perhaps it would seem more plausible if we remembered that he is talking about moral rights, not legal ones. A legal right (e.g., a ticket-collector's right to collect a fare from a train passenger) does, of course, entail a legal right to the use of powers of enforcement, since this is precisely one of the points of legalizing the right. (*Contra* Steiner, powers are not produced by rights or by entailment from rights. When Steiner says that 'having a right to A entails having a right (strictly, a power) to enforce the doing of A', and when he writes later of 'a further right (power) to force', this conflates the two distinct notions of rights and powers in a most misleading way.) What Nozick is saying, however, is not that legal rights but that moral rights carry no entitlement to the use of powers of enforcement for their maintenance. If such rights are to be maintained, then, as he says, 'Perhaps one must merely take the moral realm seriously . . . ' (p. 92). In that case, however, one wonders from whence arises the right to force people not to commit murder, which Nozick clearly posits in his first example. (It is interesting to note that, in his 1966 article on 'Coercion', reprinted in *Philosophy, Politics and Society, Series* 4, ed. P. Laslett *et al.*, Blackwell, 1972, Nozick seems less sure of the existence of a moral right to use force to defeat force, saying only that 'It is difficult to determine whether there is a presumption against threatening someone against (or coercing someone into not), e.g. murdering someone else . . . ', p. 124, note 24.) Perhaps it is this example, rather than the one which Steiner has criticized, which might reveal a flaw in Nozick's thesis.

Perhaps, though, distinctions between moral and legal rights would in any case make little difference to people's ideas about rights of enforcement. Many philosophers, it seems, intuitively think that rights entail a right of right-enforcement. For example, in a recent article in *Mind* (October, 1980), Cheyney Ryan writes: 'Your promise undoubtedly gives me the right to demand that you fulfil your promise, to threaten you if you do not, etc.' (p. 490). This is perhaps a little vague because of the 'etc.', but, from the context, it is clear that by threats Ryan means specifically threats of force (e.g., a threat to punch you in the nose, if you do not fulfil your promise). Now, like Nozick (p. 90), I find this very alarming, since it amounts to an assertion of a moral right to coerce those who are obliged to us, and thus to asserting that we have a reasonable claim to a moral right to coerce those whom we reasonably take to be obliged to us. I cannot see why I necessarily have a right to punch someone in the nose if he seems likely to break a promise to me, since there might be all sorts of good moral reasons for his contemplating the breaking of that particular promise, and, even if there were not, I can think of no reason why it should be right for me to try to force him to keep his promise by punching his nose or threatening to do so. If I said that I was going to break a promise to you, then you would have a right to demand an explanation from me and, if you were not satisfied with it, to regard me as weak or immoral or, perhaps, intellectually confused. I don't think that you'd have a right to punch me in the nose or threaten to do so in order to try to force me to keep my word.

University of London Goldsmiths' College © PATRICK WILSON 1981

STEINER ON RIGHTS AND POWERS

By David Miller

HILLEL Steiner defends the view that every right carries with it a corresponding power to enforce the obligation correlative to the right (Analysis, 41.1, January 1981, p. 50). On this view if I have a right to a piece of property (say), I necessarily have the further right to prevent another taking it, to recover it once taken, and so forth. In exercising this power I subject others to further obligations, for instance the obligation to hand back the stolen property; alternatively I may forbear to exercise my power, in which case I am waiving my primary right. Against this view Nozick has suggested that X may promise Y to do A, but before promising may extract a reciprocal promise from Y that he will not enforce X's promise, in which case Y has a right that X do A but no corresponding power to enforce. Steiner claims that this example is self-rebutting. He asks whether X's right that Y not enforce X's promise entails a further right on X's part to force Y not to enforce this promise. 'On his general thesis—that a right to something does *not* entail a right to enforce the correlative obligation—Nozick is committed to denying that X has this right. Yet if X lacks this further right, Y must lack its correlative liability which is one to be subject to an obligation not to force. But the implied effect of Y's promise is that it places him, Y, under just such an obligation. So Nozick's argument is self-contradictory since it implies at once that Y both is, and cannot be placed, under an obligation not to force.' (In Steiner's original 'X' is 'Nozick' and 'Y' is 'Hart'.)

This argument contains two errors. The first is the unwarranted shift in the first sentence of the passage quoted from 'a right does not entail a corresponding power' to 'a right entails the absence of a corresponding power'. The most plausible interpretation of Nozick's position is that a right may sometimes or even standardly carry with it a corresponding power, but that it does not *necessarily* do so. Thus a promise usually creates both a right and a power in the recipient, but in the example cited the power is removed without the primary right disappearing. So Nozick need not say that X lacks the right to force Y not to force. But even if he did say this, his view would still not be self-contradictory. For Steiner's second error is to conflate Y's *obligation* to refrain from forcing X with his *liability* to be forced to refrain. If X lacks the right to force Y, then of course Y has no correlative liability. But he still has the obligation not to force X, since this is precisely what he promised X in the first place. There is nothing incoherent here unless one supposes that Y cannot have an obligation without being liable to be forced to perform it. But that supposition is precisely equivalent to the proposition that Steiner wants to prove, namely, that every right carries with it a power to enforce.

The false claim that every right entails a corresponding power needs to be distinguished from two other claims that are more likely to be true. The first is that it is not accidental, but something like a conceptual truth, that most rights carry powers with them. In understanding what a right is, in other words, one also understands that in paradigm cases the right-holder is able to choose whether or not to enforce his right. But this is of course compatible with recognising the existence of non-paradigm cases where the power to enforce is lacking.[1] The second is that creatures incapable of exercising powers are by the same token incapable of possessing rights. To exercise a power one must be able to distinguish actions which violate one's rights from other (possibly equally unwelcome) actions which do not, and then be capable of issuing a moral or legal demand for reparation. It would be incongruous to attribute primary rights to creatures lacking these capacities, since their moral status could be described entirely in terms of the duties owed to them by others. This is why animals are properly said to lack rights.

Does anything of substance follow from this? Steiner's view entails, *inter alia*, that future generations have no rights against us, since they necessarily lack the powers to enforce that would correspond to those rights. Anyone who believes that our obligations to our descendents need to be spelt out in terms of their rights ought to be relieved that the view in question is false.

Nuffield College, Oxford © DAVID MILLER 1981

---

[1] This is, I think, the gist of Hart's position in 'Bentham on Legal Rights' in A. W. B. Simpson (ed.), *Oxford Essays in Jurisprudence*, Second Series (Clarendon Press, Oxford, 1973). But such a claim may still be too strong, as I have suggested in *Social Justice* (Clarendon Press, Oxford, 1976), pp. 62–3.

A NEGATION-FREE VERSION OF THE BERRY PARADOX

By ROBERT E. KIRK

THE Berry paradox [1], arising from the consideration of such phrases as 'the least natural number not definable in English by a phrase containing fewer than seventeen words', seems to depend essentially on the occurrence of negation in these phrases. No paradox seems to be generated, for example, by the phrases 'the least natural number definable in English by a phrase containing sixteen or fewer words' or 'the greatest natural number definable in English by a phrase containing sixteen or fewer words'. This note offers a relative of the Berry paradox based on a negation-free phrase, thus providing a companion to a result of Löb [2], which showed that negation is not essential in the derivation of paradox from the notion of truth. Löb exhibited a paradox concerning truth that did not involve negation, and it is the aim of this note to show that the same can be done for definability. I take no stand on the issue whether the paradoxes considered here are real or only apparent; I am content merely to offer a paradox, or appearance of one, that avoids negation and seems to be of the same sort as Berry's.

Consider the (incomplete) phrase 'the largest integer definable by an English phrase containing . . . or fewer words', and let the Berry function, B, be that function defined by this phrase and having the positive integers as its domain. It will be shown now that the existence of the English phrase 'one plus the largest integer definable by an English phrase containing . . . or fewer words' assures that $B(n + 14) > BB(n)$ for all positive n. To see this, suppose that E is any English phrase defining n in k words, and note that the result of filling the gap in the last phrase with E is a phrase defining $B(n) + 1$ in $k + 14$ words. But, since $B(k + 14)$ is the largest integer definable in $k + 14$ or fewer words, $B(k + 14) \geqslant B(n) + 1$. Finally, as there must be a phrase E_n defining B(n) in n or fewer words, it may be concluded that $B(n + 14) > BB(n)$ for all n. Observe now that the compactness of English expressions denoting numbers guarantees that there must be an integer N such that $N + 14 \leqslant B(N)$. In fact, $N = 4$ is one such, since 'the square of five' is a four-word phrase defining 25, and $4 + 14 < 25 \leqslant B(4)$. Finally, since B is monotone ($B(n) \leqslant B(m)$ whenever $n \leqslant m$), $B(N + 14) \leqslant BB(N)$, a contradiction.

The argument I have given is modelled after one used by Rado [3] to show that certain functions are not recursive.

University of Washington

© Robert E. Kirk 1981

REFERENCES

[1] Russell, B. 'Les Paradoxes de la Logique', Revue de Métaphysique et de Morale, 14 (1906), pp. 627–650.
[2] Löb, M. H. 'Solution of a Problem of Leon Henkin', Journal of Symbolic Logic, 20 (1955), pp. 115–118.
[3] Rado, T. 'On Non-Computable Functions', Bell System Technical Journal, XLI (1962), pp. 877–884.

RATIO edited by MARTIN HOLLIS

Volume XXIII *June 1981* *No. 1*

CONTENTS

PETER ALEXANDER	*Locke on Substance-in-General II*
D.H. MELLOR	*'Thank Goodness That's Over'*
EMILY GROSHOLZ	*Wittgenstein and the Correlation of Logic and Arithmetic*
J.M. HINTON	*The Fountain of Truth: A Eulogy of Ramsey*
GEORG BRAUNGART	*Ethics and its Amoral Justification in Hobbes*
JONATHAN HARRISON	*Three Philosophical Fairy Stories*
Reviews	
ALEC FISHER	*Wittgenstein on the Foundations of Mathematics*, by Crispin Wright
DAVID-HILLEL RUBEN	*Main Currents of Marxism*, by Leszek Kolakowski

Subscriptions
Institutions: £18.50 (UK) £22.75 (overseas) $49.50 (USA) $57.00 (Canada)
Individuals: £12.50 (UK) £15.25 (overseas) $32.50 (USA) $37.25 (Canada)

Orders to: Journals Department, Basil Blackwell Publisher Ltd.,
108 Cowley Road, Oxford OX4 1JF, England.

The German edition of *Ratio* is published by FELIX MEINER VERLAG,
Richardstrasse 47, 2000 Hamburg 76, West Germany.

Published by BASIL BLACKWELL, Oxford

THEORIA

A Swedish journal of philosophy
Edited by Peter Gärdenfors, Bengt Hansson
and Krister Segerberg

Volume 46, part 2–3

Alan Musgrave Wittgensteinian instrumentalism
Martin M. Tweedale Abailard and Ockham: Contrasting defences of nominalism
M. J. Cresswell Jackson on perception
Gregory Currie and **Graham Oddie** Changing numbers
Pavel Tichý The transiency of truth
Krister Segerberg A note on the logic of elsewhere
Review by **Wilf Malcolm**

THEORIA, which invites contributions in all fields of analytic philosophy, usually publishes three issues a year. Subscriptions and requests for back numbers should be sent to *Editor*, THEORIA, *Filosofiska institutionen, Kungshuset i Lundagård, S-223 50 Lund, Sweden.* Subscription for one year is SwCr 60:– or U.S. $ 16.–, for three years SwCr 160:– or U.S. $ 40.–.

JOURNAL FOR THE THEORY OF SOCIAL BEHAVIOUR

General Editor: P. F. SECORD Co-Editor: P. COLLETT
Associate Editors: A. GIDDENS, P.MANICAS, C. SMITH
Managing Editor: J. S. WEST

Volume 11, No. 2 **July 1981**

A. W. McHOUL: Ethnomethodology and the Position of Relativist Discourse

WILLIAM H. FODDY: Obscenity Reactions: Toward a Symbolic Interactionist Explanation

JOSEPH LYONS: Memory traces and infantile amnesia: A reconsideration of the work of Erwin Straus

JOSEPH LYONS: Why Human Development Requires More than One Mode of Experience

PHILIP L. KNOWLES: Ecological Perspective Applied to Social Perception

MAURY SILVER and JOHN SABINI: Procrastinating

Annual subscriptions: Institutions £20.50 (inland) £25.25 (overseas) $52.50 (USA) $60.00 (Canada); Individuals £10.50 (inland) £13.00 (overseas) $30.00 (USA) $34.50 (Canada)

Orders to: JOURNALS DEPARTMENT, **BASIL BLACKWELL PUBLISHER.**,
108 Cowley Road, Oxford OX4 1JF, England.

The Philosophical Quarterly

Special Numbers:
The Editors intend to devote an issue in 1982 to articles on **Scientific Realism** and one issue in 1983 to articles on **Medical Ethics.** Contributions should be submitted by **31 December of the preceding year,** to the Executive Editor, The Philosophical Quarterly, The University, St Andrews, Scotland.

Vol. 31 *No.* 124 **Special Issue on Epistemology** *October* 1981

Psychological Verbs and Referential Attitudes — *Lauchlan Chipman*
The Community of Knowledge — *Michael Welbourne*
Justified True Belief is Knowledge — *Graham Dawson*
Indirect Perception and Sense Data — *E. J. Lowe*
DISCUSSIONS
Empiricism and the Private Language Argument — *Kim Davies*
Descartes, Kant and Self-Consciousness — *Stephen Priest*
CRITICAL STUDY
C. Wright, Wittgenstein on the Foundations of Mathematics — *Cora Diamond*
BOOK REVIEWS — BOOKS RECEIVED

Annual Subscriptions:
Individuals £6.00 ($15.00) Institutions £12.00 ($30.00)
for orders placed directly with the Journals Manager, Scottish Academic Press Ltd., 33 Montgomery Street, EDINBURGH EH7 5JX, Scotland.